

REPOSITORIO ACADÉMICO UPC

Reconocimiento de gestos estáticos del abecedario de la lengua de señas peruana utilizando cámaras de baja resolución

Item Type	info:eu-repo/semantics/bachelorThesis
Authors	Berru Novoa, Bryan Jos
Publisher	Universidad Peruana de Ciencias Aplicadas (UPC)
Rights	info:eu-repo/semantics/openAccess
Download date	07/10/2024 15:32:48
Item License	http://creativecommons.org/licenses/by-nc-sa/4.0/
Link to Item	http://hdl.handle.net/10757/628221



UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

FACULTAD DE INGENIERÍA

PROGRAMA ACADÉMICO DE CIENCIAS DE LA COMPUTACIÓN

Reconocimiento de gestos estáticos del abecedario de la lengua de señas
peruana utilizando cámaras de baja resolución

TESIS

Para optar el título profesional de Licenciado en Ciencias de la Computación

AUTOR(ES)

Berru Novoa, Bryan Jos (0000-0003-0565-1446)

ASESOR

González Valenzuela, Ricardo Eugenio (0000-0002-6727-2516)

Lima, 20 de Noviembre de 2019

DEDICATORIA

Esta va tesis va dedicada primeramente a mis padres que me alentaron día a día a seguir adelante a no rendirme a pesar de las dificultades que pase. A mis asesores, profesores y amigos por la paciencia y todos los consejos que me dieron a lo largo del proyecto, y, sobre todo, gracias a Dios por darme las fuerzas día a día y que siempre fue mi fuente de esperanza en todos los tiempos de dificultad.

AGRADECIMIENTOS

Se agradece el apoyo brindado por la asociación ENSEÑAS PERÚ, por habernos apoyado brindando expertos en la lengua de señas peruana, que nos ayudaron a reproducir y evaluar el dataset generado en esta tesis.

RESUMEN

El reconocimiento de gestos del lenguaje de señas a través del procesamiento de imágenes y el aprendizaje automático han sido ampliamente estudiados en años recientes. En esta tesis se presenta un conjunto de datos que consta de 2400 imágenes de los gestos estáticos del lenguaje de señas peruano alfabeto, además de aplicarlo a un sistema de reconocimiento de gestos con cámaras de baja resolución. Para el reconocimiento de gestos, se utilizó el descriptor de características de Histograma de Gradientes Orientados, junto con 4 algoritmos de clasificación. Los resultados mostraron que utilizando el Histograma de Gradientes Orientados junto con el algoritmo de clasificación de Support Vector Machine, se obtuvo el mejor resultado con un accuracy de 89.46% y el sistema pudo reconocer los gestos con variaciones de traslación, rotación y escala.

Palabras clave: Procesamiento de imágenes; Histograma de Gradientes Orientados; Reconocimiento de lenguaje de señas; Aprendizaje automático.

Recognition of static gestures of the Peruvian sign language alphabet using low resolution cameras

ABSTRACT

The recognition of sign language gesture through image processing and Machine Learning has been widely studied in recent years. This article presents a dataset consisting of 2400 images of the static gestures of the Peruvian sign language alphabet, in addition to applying it to a recognition system of gestures using low resolution cameras. For the gesture recognition, the Histogram Oriented Gradient feature descriptor was used, along with 4 classification algorithms. The results showed that Histogram Oriented Gradient, along with Support Vector Machine, got the best result with an 89.46% accuracy and the system was able to recognize the gestures with variations of translation, rotation and scale.

Keywords: Image Processing; Machine Learning; Histogram Oriented Gradient; Sign Language Recognition

TABLA DE CONTENIDOS

CONTENIDO

1	INTRODUCCIÓN	1
2	DESCRIPCIÓN DEL PROYECTO:	1
2.1	OBJETO DE ESTUDIO	1
2.2	DOMINIO DEL PROBLEMA	2
2.3	OBJETIVOS DE LA INVESTIGACIÓN	2
2.3.1	Objetivo General	2
2.3.2	Objetivos Específicos	2
2.4	INDICADORES DE ÉXITO	3
2.5	PLANIFICACIÓN DEL PROYECTO	3
2.5.1	Entregables	3
2.5.2	Exclusiones	4
2.6	PLAN DE GESTIÓN DE TIEMPO	4
2.7	RESTRICCIONES	5
2.8	PLAN DE GESTIÓN DE RECURSOS HUMANOS	5
3	ACHIEVEMENTS OF THE STUDENT OUTCOMES	7
3.1	STUDENT OUTCOME A: APLICACIÓN DE CONOCIMIENTOS DE CIENCIAS	7
3.2	STUDENT OUTCOME B: PLANIFICACIÓN Y CONDUCCIÓN DE EXPERIMENTOS ...	7
3.3	STUDENT OUTCOME C: DISEÑO DEL MÉTODO COMPUTACIONAL	8
3.4	STUDENT OUTCOME D: TRABAJOS EN EQUIPOS MULTIDISCIPLINARIOS	9
3.5	STUDENT OUTCOME E: RESPONSABILIDAD ÉTICA Y PROFESIONAL	9
3.6	STUDENT OUTCOME F: COMUNICACIÓN EFECTIVA	11
3.7	STUDENT OUTCOME G: ANÁLISIS DEL IMPACTO DE LA SOLUCIÓN DE INGENIERÍA	11
3.8	STUDENT OUTCOME H: NECESIDAD DE APRENDIZAJE DE POR VIDA	12
3.9	STUDENT OUTCOME I: USO DE TÉCNICAS Y HERRAMIENTAS	13
3.10	STUDENT OUTCOME J: APLICACIÓN DE FUNDAMENTOS MATEMÁTICOS	14
3.11	STUDENT OUTCOME K: DISEÑO Y CONSTRUCCIÓN DE SISTEMAS COMPLEJOS	15
4	ESTADO DE ARTE	17

4.1	MÉTODO:	17
4.2	RESUMEN POR ARTÍCULO	18
4.2.1	Segmentación de manos con cámaras de baja resolución	18
4.2.2	Extracción de características de gestos de mano estáticos mediante descriptores de imagen	23
4.2.3	Algoritmos de clasificación aplicados al reconocimiento de gestos de mano estáticos	27
4.3	CONCLUSIONES	33
4.3.1	Conclusiones para la tipología de segmentación de manos con cámaras de baja resolución	33
4.3.2	Conclusiones para la tipología de extracción de características de gestos de mano estáticos mediante descriptores de imágenes	34
4.3.3	Conclusiones para la tipología de algoritmos de clasificación aplicados al reconocimiento de gestos de mano estáticos	34
5	MARCO TEÓRICO	35
5.1	RECONOCIMIENTO DE GESTOS DE MANO	35
5.2	RECONOCIMIENTO BASADO EN VISIÓN COMPUTACIONAL	35
5.3	PROCESAMIENTO DE IMÁGENES	35
5.4	DESCRIPTORES DE IMAGEN	36
5.5	HISTOGRAM ORIENTED GRADIENT	37
5.6	LENGUA DE SEÑAS	38
5.7	MACHINE LEARNING	39
5.8	SUPPORT VECTOR MACHINE	39
6	DESARROLLO DEL PROYECTO	40
6.1	SISTEMA PROPUESTO	40
6.2	GENERACIÓN DEL DATASET	42
6.3	HADWARE Y SOFTWARE	43
6.4	DATASET DE IMÁGENES	43
6.5	DATASET DE CARACTERÍSTICAS	44
6.6	ENTRENAMIENTO Y VALIDACIÓN	44
7	RESULTADOS DEL PROYECTO	45

7.1.1	Support Vector Machine.....	45
7.1.2	K-Nearest Neighbors	46
7.1.3	MultiLayer Perceptron	48
7.1.4	Random Forest.....	49
7.1.5	Conclusiones.....	50
7.2	APLICACIÓN DEL MODELO DE RECONOCIMIENTO DE GESTOS	51
8	GESTIÓN DEL PROYECTO	53
8.1	GESTIÓN DE TIEMPO	53
8.2	GESTIÓN DE LAS COMUNICACIONES	56
8.3	GESTIÓN DE LOS RIESGOS	57
8.4	LECCIONES APRENDIDAS	58
9	CONCLUSIONES	59
10	RECOMENDACIONES	61
11	GLOSARIO.....	62
12	REFERENCIAS	63
13	ANEXOS	65

ÍNDICE DE TABLAS

Tabla 1: Problemas y causas del proyecto	2
Tabla 2: Restricciones del proyecto	5
Tabla 3: Roles y Responsabilidades de los interesados del proyecto	5
Tabla 4: Recopilatorio de artículos del estado del arte	18
Tabla 5: Porcentaje de acierto de la detección de los gestos	22
Tabla 6: Resultados de la experimentación	23
Tabla 7: Porcentaje de reconocimiento del sistema utilizando distancia Euclidiana y Mahalanobis	25
Tabla 8: Resultados del Sistema A-ASLR	26
Tabla 9: Accuracy obtenido	27
Tabla 10: Resultados del reconocimiento de 10 gestos para 5 usuarios	28
Tabla 11: Accuracy de las arquitecturas	29
Tabla 12: Duración del epoch de las arquitecturas de CNN (con GPU)	30
Tabla 13: Parámetros del HOG	44
Tabla 14: Mejores resultados obtenidos por cada algoritmo en términos de accuracy	50
Tabla 15: Cronograma del Proyecto para el 1º ciclo académico	54
Tabla 16: Cronograma del proyecto para el 2º ciclo académico	55
Tabla 17: Matriz de Riesgos	57

ÍNDICE DE FIGURAS

Figura 1: Sprint 1- Planificación	4
Figura 2: Sprint 2 – Iteración 1	4
Figura 3: Sprint 3 – Iteración 2	5
Figura 4: Sistema de reconocimiento de gestos propuesto	16
Figura 5: Gestos reconocidos por el sistema	19
Figura 6: Resultados del desempeño del sistema	19
Figura 7: Gestos reconocidos por el sistema	20
Figura 8: Funciones del mouse	21
Figura 9: Reconocimiento de la mano mediante la técnica propuesta	22
Figura 10: Lenguaje de señas Vietnamita	24
Figura 11; Distancia entre los centroides y el borde de la mano	27
Figura 12: Gestos a reconocer por el sistema	28
Figura 13: Arquitectura de la red MAdaline	31
Figura 14: Comparación de MAdaline con otros algoritmos	31
Figura 15: Accuracy del MLP en base al n° de celdas usadas en el HOG	32
Figura 16: Accuracy del Sparse Autoencoder en base al n° de parámetros	33
Figura 17: Filtros de pre-procesamiento de mano	36
Figura 18: Descriptor HOG	38
Figura 19: Abecedario del lenguaje de señas peruano	38
Figura 20: Plano de separación	39
Figura 21: Fases del sistema de reconocimiento de gestos propuesto	40
Figura 22: Metodología para generar el dataset	42
Figura 23: Detección, recorte y redimensión de la región de la mano	43
Figura 24: Resultados de Support Vector Machine utilizando HOGS de 4 a 9 bins	45
Figura 25: Matriz de confusión de SVM con un HOG de 9 bins	46
Figura 26: Resultados de K-Nearest Neighbors utilizando HOGS de 4 a 9 bins	46
Figura 27: Matriz de confusión de KNN con un HOG de 9 bins	47
Figura 28: Resultados de MultiLayer Perceptron utilizando HOGS de 4 a 9 bins	48
Figura 29: Matriz de confusion de MultiLayer Perceptron utilizando un HOG de 8 bins	49
Figura 30: Resultados de Random Forest utilizando HOGS de 4 a 9 bins	49
Figura 31: Matriz de confusión de Random Forest utilizando un HOG de 6 bins	50

Figura 32: Reconocimiento de la letra “a” utilizando el modelo de clasificación de SVM	
.....	51

1 INTRODUCCIÓN

La lengua de señas es utilizada por personas que presentan alguna dificultad para hablar u oír, y les permite comunicarse con otras personas que pueden o no presentar las mismas dificultades. No obstante, este lenguaje no es universal, por lo que cada país cuenta con una lengua de señas propia. Diversos estudios dentro del área de procesamiento de imágenes han creado sus propios datasets que incluyen gestos de manos de las lenguas de señas de diversos países (americano, iraní, egipcio, chino, etc.). Estos datasets fueron utilizados para diversos fines tales como: la interacción humano-computador, para facilitar la comunicación de personas sordas, la manipulación de robots utilizando este tipo de lengua, o el desarrollo de programas educativos que permitan mejorar la enseñanza de este tipo de lenguajes en sus países. No obstante, en nuestra etapa de investigación, no se ha encontrado una fuente que haya utilizado un dataset de la lengua de señas peruana, por lo que siendo esta nuestra principal motivación, nos proponemos a generar un modelo de reconocimiento de gestos de mano, con el fin de clasificar 24 gestos estáticos del abecedario de la lengua de señas peruana, con el fin de aplicarlo dentro de un sistema de reconocimiento que permita facilitar la enseñanza del idioma español en las escuelas de sordos.

Este proyecto tiene por finalidad presentar la información requerida para el desarrollo de este modelo de reconocimiento de gestos aplicado, únicamente, a cámaras de baja resolución. Además, se presentará el modelo dentro de un sistema de reconocimiento como método de validación comparando el accuracy con 4 algoritmos de clasificación.

La gestión del proyecto de investigación será en base a los pasos de la metodología PMBOK que son: inicio, planificación, ejecución y cierre, los cuales abarcarán la definición del comienzo y fin del proyecto, definiendo el trabajo desarrollado a través de las fases mencionadas y a los involucrados en cada una de estas.

2 DESCRIPCIÓN DEL PROYECTO:

En el presente capítulo se realizará la descripción a detalle del proyecto, basándonos en el PMBOK. Esta sección consta de diversas subsecciones enfocadas a la gestión del proyecto, desde los objetivos al manejo de los recursos y tiempo del mismo.

2.1 Objeto de Estudio

Este proyecto tiene como fin: Aplicar un modelo de clasificación de reconocimiento de gestos en tiempo real del abecedario de lengua de señas peruana dentro de un sistema de

reconocimiento que, utilizando una cámara de baja resolución, permita enseñar a niños sordos y no sordos acerca del idioma español. Para lograr esto, se utilizará un descriptor de imagen, con el fin de analizar las características de esta imagen y, obtener un vector descriptor que servirá para entrenar el algoritmo de clasificación. La metodología que se aplicara será la siguiente: Primero, se analizara los mejores algoritmos de clasificación de lenguaje de señas, en términos de accuracy. Una vez obtenido el mejor algoritmo para el modelo, se procederá a diseñar el diagrama de flujo de la detección de gestos en tiempo real, para el reconocimiento de señas. Después, se construirá el dataset de la lengua de señas peruana y se generará el modelo con la metodología propuesta. Finalmente, se integrará el modelo dentro de un sistema de reconocimiento con el fin de validar la eficiencia de este.

2.2 Dominio del Problema

Tabla 1: Problemas y causas del proyecto

Problema	Causas
El reconocimiento en tiempo real de la lengua de señas de diversos países mediante cámaras de baja resolución, utilizando Machine Learning y procesamiento de imágenes.	<ul style="list-style-type: none"> Diversos métodos de detección de silueta de manos utilizando visión computacional¹ Modelos de clasificación de gestos del lenguaje de señas enfocados en precisión² Amplios usos en diversas áreas para la detección de gestos³

Fuente: Elaboración Propia

2.3 Objetivos de la Investigación

2.3.1 Objetivo General

Aplicar un modelo de clasificación de reconocimiento de gestos estáticos del abecedario de la lengua de señas peruana, utilizando cámaras de baja resolución, que presente un accuracy mayor a 90%, comparándolo con los algoritmos de:

- Support Vector Machine
- K-Nearest Neighbors
- MultiLayer Perceptron
- Random Forest

2.3.2 Objetivos Específicos

- OE1: Analizar los mejores algoritmos de clasificación para el reconocimiento de gestos de mano estáticos en términos de precisión.

- OE2: Diseñar un proceso de extracción de características de la imagen y clasificación de gestos, para el reconocimiento de señas.
- OE3: Construir un dataset de imágenes del abecedario de la lengua de señas peruana para los 24 gestos estáticos a reconocer.
- OE4: Desarrollar un modelo de clasificación eficiente, en términos de accuracy, para el reconocimiento de 24 gestos estáticos del abecedario de la lengua de señas peruana.

2.4 Indicadores de Éxito

- Para el OE1:
 - IE1: Documento de análisis comparativo de los algoritmos de clasificación, basado en trabajos relacionados al reconocimiento de gestos de mano, además de pruebas de concepto, aprobado por el cliente.
- Para el OE2:
 - IE2: Documento del diseño del proceso de extracción de características de la imagen y clasificación de gestos, a través de un diagrama de flujo, aprobada por el cliente
- Para el OE3:
 - IE3: Dataset que contenga los gestos estáticos del abecedario de la lengua de señas peruana aprobado por el cliente y la organización ENSEÑAS PERÚ.
- Para el OE4:
 - IE4: Documento del desarrollo del modelo de reconocimiento de gestos con diversos algoritmos de clasificación, aprobado por un especialista en Machine Learning.

2.5 Planificación del Proyecto

2.5.1 Entregables

El alcance del proyecto incluirá:

- Un dataset de imágenes de los 24 gestos estáticos del abecedario de la lengua de señas peruana.
- Un modelo de clasificación para el reconocimiento de los 24 gestos estáticos del abecedario de la lengua de señas peruana en tiempo real.

- Un software en desktop que aplique el modelo de clasificación propuesto.

2.5.2 Exclusiones

- La detección de gestos solo se limita al abecedario de la lengua de señas peruana (no incluye números)
- Las letras a reconocer se conforman solo por gestos estáticos (no se incluye LL, J, Z, Ñ)
- El programa solo podrá ser ejecutado en el sistema operativo Windows

2.6 Plan de Gestión de Tiempo

Sprint 1:

Cronograma: Planificación

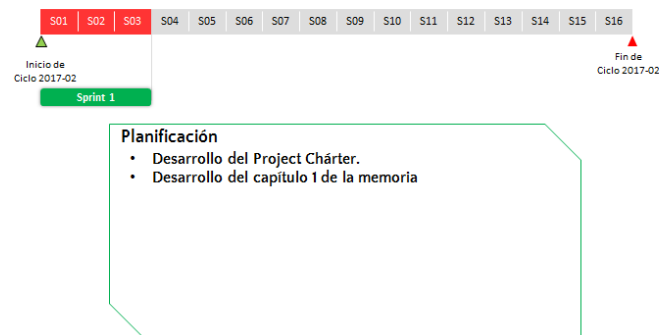


Figura 1: Sprint 1- Planificación

Fuente: Elaboración propia

Sprint 2:

Cronograma: Iteraciones

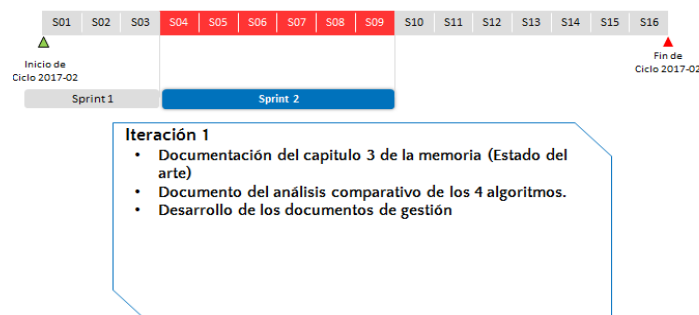


Figura 2: Sprint 2 – Iteración 1

Fuente: Elaboración propia

Sprint 3:

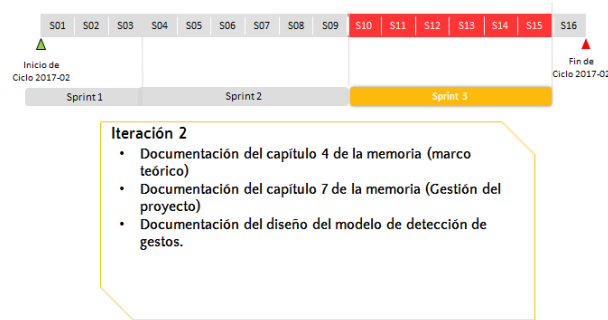


Figura 3: Sprint 3 – Iteración 2

Fuente: Elaboración propia

2.7 Restricciones

Tabla 2: Restricciones del proyecto

Restricciones	Descripción
Restricción 1	El proyecto durara un año académico
Restricción 2	Plazos de entrega impostergables

Fuente: Elaboración propia

2.8 Plan de Gestión de Recursos Humanos

Tabla 3: Roles y Responsabilidades de los interesados del proyecto

Responsable	Roles	Responsabilidades
Rosario Villalta Luis Vives Pedro Shiguihara Jimmy Armas	Comité de Proyectos	Evaluar el proyecto bajo los estándares establecidos por la carrera de Ciencias de la Computación e Ingeniería de Software de la Universidad Peruana de Ciencias Aplicadas
Ricardo González	Cliente del proyecto	-Monitorea y valida los avances del proyecto
Bryan Berru	Jefe del proyecto	-Planifica, dirige y cumplen los objetivos planteados.
Daniel Burga	Gerente del proyecto	-Monitorea y valida los planes de gestión y de solicitud de recursos del proyecto

Marita Padilla	Directora Ejecutiva de ENSEÑAS PERÚ	Brinda un video con las señas requeridas para este proyecto
----------------	--	---

Fuente: Elaboración Propia

3 ACHIEVEMENTS OF THE STUDENT OUTCOMES

En este capítulo se verá la relación que el proyecto posee con los Student Outcomes planteados por la UPC que todo alumno de la carrera de Ciencias de la Computación debe tener. Estos se dividen en 11 criterios, entre los cuales se resaltan los conocimientos adquiridos a lo largo de toda la carrera.

3.1 Student Outcome A: Aplicación de Conocimientos de Ciencias

Para el reconocimiento de gestos de mano, se requirió aplicar los fundamentos de 2 áreas de Ciencias de la computación: procesamiento de imágenes e inteligencia artificial (específicamente el área de Machine Learning). Para el primero, se aplicó una técnica extracción de fondo utilizando el algoritmo de “Gaussian Mixture-based Background Segmentation Algorithm” (Zoran Z., et. al., 2006), además del descriptor de imagen de Histogram Oriented Gradient (N. Dalal et. al., 2005) para generar el dataset de características. Para el segundo, se experimentó y analizó diversos modelos de clasificación, los cuales fueron generados a partir de 4 algoritmos de clasificación, con el fin de medir la eficiencia de cada uno de los modelos en términos de accuracy.

La evidencia del cumplimiento de este Student Outcome se encuentra en los capítulos 6 y 7 de la memoria del proyecto.

En el capítulo 6, sección 6.1 se explica el proceso que utiliza el algoritmo de HOG para obtener las características de la mano. La sección 6.2, explica la metodología que se aplicó para la recolección de las imágenes y la generación del dataset. Adicionalmente, la sección 6.3, explica los parámetros utilizados para conseguir las características y la fórmula de la métrica a utilizar para la experimentación.

3.2 Student Outcome B: Planificación y Conducción de Experimentos

Para el planteamiento de la problemática y los objetivos del proyecto, se realizó un análisis del estado del arte para identificar el problema, revisar las soluciones dadas a ese problema y el aporte innovador que se podría ofrecer. Se identificó la falta de un dataset de gestos estáticos del abecedario de la lengua de señas peruana, el cual permitiría la posibilidad de crear un programa de reconocimiento de señas que permita facilitar la enseñanza de este lenguaje a las personas. La evidencia de esto se encuentra en el capítulo 2 y 4 de la memoria del proyecto.

La planificación y gestión del proyecto se desarrolló utilizando el PMBOK. La evidencia de esto se encuentra en el capítulo 8 de la memoria del proyecto.

En cuanto a los experimentos, el objetivo principal fue obtener el mejor modelo de clasificación para los 24 gestos estáticos del abecedario de la lengua de señas peruana, teniendo como métrica principal el accuracy, debido a que el dataset cuenta con el mismo número de instancias para cada gesto; además de que este modelo obtenga un accuracy mayor a 90% utilizando la técnica de cross-validation (15 folds).

En la conducción de estos experimentos, se recolecto un dataset de 3600 imágenes donde cada gesto contenía un total de 150 imágenes. Estas imágenes pasaron por el método de preprocesamiento propuesto (extracción de fondo y HOG), con el fin de obtener un vector de características que permita ser el dato de entrada para los 4 algoritmos de clasificación seleccionados en el estado del arte, los cuales son:

- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- MultiLayer Perceptron (MLP)
- Random Forest (RF)

La evidencia de los experimentos realizados se encuentra detallado en los capítulos 6 y 7 de la memoria del proyecto. Asimismo, todos los experimentos fueron validados por un especialista en Machine Learning.

Por otro lado, el dataset fue validado por un especialista en la lengua de señas peruana, cuya evidencia se encuentra en las actas anexadas en el proyecto.

3.3 Student Outcome C: Diseño del Método Computacional

El diseño del programa se basó en estudios previos de reconocimiento de gestos de mano estáticos. El proceso para armar este sistema de reconocimiento de gestos consta de 3 áreas: La primera se enfoca en la recolección de las imágenes y la generación del dataset de características. La segunda se enfoca en utilizar este dataset de características dentro de algoritmos de clasificación, utilizando scikit-learn, con el fin de generar un modelo de clasificación. Finalmente, la tercera se enfoca en el proceso realizado para reconocer un gesto realizado por el usuario frente a la cámara, utilizando el modelo entrenado.

Cada uno de estos pasos fue detallado en el objetivo específico 2. Además de ello, el capítulo 6 de la memoria del proyecto, muestra los diagramas de las actividades que se realizaron en cada una de las 3 áreas mencionadas.

3.4 Student Outcome D: Trabajos en Equipos Multidisciplinarios

Durante el desarrollo del proyecto, se desempeñó el rol de jefe de proyecto y desarrollo. En las actividades realizadas, se ha requerido la participación de la asociación ENSEÑAS PERU con el fin de construir el dataset de gestos estáticos del abecedario de la lengua de señas peruana. Para ello, la asociación proveyó de personas con dificultades auditivas, con las cuales se trabajó conjuntamente para obtener cada una de las imágenes del objetivo específico 3, lo cual involucro la grabación de cada uno de los gestos y la validación del dataset. Además de ello, este proyecto tuvo seguimiento por parte de especialistas en el área de gestión de proyectos (Profesor gerente) y especialistas en el área de procesamiento de imágenes y Machine Learning (Profesor cliente), con el fin de lograr un proyecto de alto estándar. La evidencia de esto se encuentra en las actas de reunión, firmadas por el cliente, gerente y la asociación ENSEÑAS PERU.

3.5 Student Outcome E: Responsabilidad Ética y Profesional

Dentro del aspecto de la responsabilidad ética, el proyecto se basó en los 10 acuerdos establecidos en el código ético de la IEEE. Con ello, se aseguró un desarrollo integro del proyecto, además de un buen trato profesional con todos los miembros involucrados.

Como evidencia de ello se adjuntan las actas de validación que determinan el cumplimiento de cada uno de los objetivos planteados a lo largo del proyecto, donde se evidencia el cumplimiento de cada etapa, con la validación de un experto en el área. Adicionalmente, para la elaboración del artículo científico al igual que el estado del arte y toda la información empleada para la obtención del resultado final, todos los autores fueron citados apropiadamente de modo que no se sostienen faltas hacia el uso indebido o robo de conocimiento.

Por otro lado, debido a que el proyecto se enfoca en reconocer los gestos estáticos del abecedario de la lengua de señas peruana, se requirió trabajar junto a la asociación ENSEÑAS PERU, los cuales son especialistas en trabajar con gente sorda y nos permitieron grabar a estas personas realizando cada uno de los 24 gestos estáticos, además de validar el dataset que se utilizaría para generar el modelo de clasificación. Esta validación se puede

encontrar en el acta de validación del objetivo específico 3 del proyecto. Además, también se incluyen las siguientes fotos con el especialista y el grupo de niños con el que se trabajó.



Dentro del aspecto de la responsabilidad profesional. Se propuso que el dataset generado en el proyecto sea distribuido abiertamente, con el fin de brindar un dataset del abecedario de la lengua de señas peruana a futuros proyectos relacionados con el reconocimiento de gestos usando este lenguaje, de manera que esos proyectos puedan utilizar este dataset como pruebas de concepto para sus algoritmos o como medio de comparación a futuros datasets que se propongan.

El dataset se puede encontrar en el siguiente link: <https://github.com/Expo99/Static-Hand-Gestures-of-the-Peruvian-Sign-Language-Alphabet> , el cual se encuentra bajo la licencia de “Creative Commons Attribution Share Alike 4.0 International”, que establece el uso personal, modificación y distribución del dataset bajo las siguientes condiciones:

- Se debe hacer referencia al creador del dataset dentro del proyecto que se utilice, además de un link a la fuente
- Todo cambio hecho en el dataset que desea ser distribuido a otras personas, debe ser hecho bajo la misma licencia

Por otro lado, el código fuente solo será distribuido a personas del ámbito científico que estén desarrollando proyectos relacionados a el reconocimiento de señas, con el fin de mantener el código dentro del uso del área científica.

3.6 Student Outcome F: Comunicación Efectiva

Para el desarrollo del proyecto se utilizó principalmente el correo electrónico, provisto por la UPC, para cualquier toma de decisiones entre el cliente y el gerente. Adicionalmente, todas las actividades que se conversaron y se realizaron en las reuniones presenciales y virtuales que se realizó con el cliente fueron detalladas en actas de reunión. Por otro lado, se mantuvo una constante comunicación con la asociación ENSEÑAS PERÚ, con el fin de acordar las reuniones para las grabaciones de cada uno de los gestos de la lengua de señas peruana. Finalmente, este proyecto fue aceptado y expuesto en el: “IEEE XXV International Conference on Electronics, Electrical Engineering and Computing” (INTERCON) realizado en el 2018.

3.7 Student Outcome G: Análisis del Impacto de la Solución de Ingeniería

La finalidad del proyecto fue el de generar un modelo de clasificación que sea capaz de clasificar 24 gestos estáticos del abecedario de la lengua de señas peruana. Este modelo fue aplicado dentro de un sistema de reconocimiento de gestos, el cual permite a los usuarios deletrear palabras y, de manera intuitiva, aprender la lengua de señas peruana.

Gracias a este proyecto se puede abrir las puertas a una nueva forma de enseñanza para la lengua de señas peruana. Un ejemplo de ello es la posibilidad de crear una página web que contenga este sistema integrado y donde un profesor pueda dar una clase virtual y, mientras les enseña a los alumnos acerca del abecedario, ellos pueden utilizar el programa para familiarizarse con los gestos que está realizando el profesor.

Adicionalmente, este proyecto fue presentado en la feria de carreras dada por la Universidad Peruana de Ciencias Aplicadas.



3.8 Student Outcome H: Necesidad de Aprendizaje de por Vida

Para este tipo de proyectos, es necesario realizar investigaciones continuas sobre herramientas, algoritmos, y modelos de diversas fuentes, las cuales fueron de suma importancia para llegar al objetivo final dada la rápida evolución de la tecnología y las ciencias. Dentro de estas investigaciones se encontró artículos científicos que aplicaban técnicas de procesamiento de imágenes y Machine Learning, con el fin de reconocer los gestos que realizaba una mano frente a una cámara de baja resolución (webcam). Dentro de esta bibliografía se analizaron estudios como el de “Real time Finger Tracking and Contour Detection for Gesture Recognition using OpenCV” (R. Gurav, K. Kadbe, 2015), el cual menciona técnicas de preprocesamiento, con el fin de reconocer la región de la mano dentro de un fondo estático. Otro estudio como el de “Sign Language Conversion Tool (SLCTool) Between 30 World Sign Languages” (Sastry A.S.C.S, et. al., 2018), menciona el uso de

descriptores de imagen como el HOG y compara su eficiencia con otros descriptores de imagen utilizando, como el algoritmo de clasificación, un SVM. Adicionalmente, otros estudios utilizan descriptores de imagen, pero con un algoritmo de clasificación distinto como el caso de “Hand gesture recognition using neural network based techniques” (Vladislava B., et. al., 2016), el cual utilizó un MLP y el estudio de “SIFT based approach on Bangla Sign Language Recognition” (Farhad Y., et. al, 2015), donde utilizó un KNN.

Por otro lado, en el transcurso del proyecto se encontró la necesidad de trabajar con una asociación externa para la creación del dataset, para lo cual se contactó con la ONG ENSEÑAS PERU, con el fin de obtener las imágenes para cada uno de los gestos del abecedario de la lengua de señas peruana.

3.9 Student Outcome I: Uso de Técnicas y Herramientas

Dentro del desarrollo del proyecto, se utilizaron 2 técnicas de preprocesamiento con el fin de obtener un gesto realizado de una mano:

- **Background Sustraction:** Para esta técnica se utilizó un modelo Gausseano para extraer el fondo de una imagen, con el fin de resaltar elementos nuevos que se coloquen dentro de la imagen. Gracias a ello, podemos extraer los elementos del fondo, colocar nuestra mano dentro de un fondo vacío (en este caso de color negro) y, cuando se quiera hacer el reconocimiento, lo único que tendrá la imagen es el gesto realizado por la mano, lo cual evita que se confunda con otros elementos del fondo.
- **Histogram of Oriented Gradients:** La técnica de HOG se utilizó con el fin de extraer un vector de características de la imagen. Este vector describe la intensidad de los gradientes de esta imagen, y sirve como dato de entrada para los algoritmos de clasificación que se utilizaron para los experimentos.

En cuanto a las herramientas utilizadas, se requirió el uso de 2 bibliotecas, con el fin de lograr el reconocimiento de gestos de mano. Estas librerías fueron:

- **OpenCV:** Una librería de procesamiento de imágenes que facilitó la implementación del HOG, así como el reconocimiento del área de la región de la mano, preprocesamiento, uso de la cámara y manejo de ventanas del programa.

- **scikit-learn:** Una librería de Machine Learning, la cual facilitó la experimentación de los algoritmos de clasificación y el análisis de los modelos generados.

3.10 Student Outcome J: Aplicación de Fundamentos Matemáticos

Para el reconocimiento de gestos de mano, se requiere entender los fundamentos matemáticos de 2 áreas de la ciencia de la computación: Procesamiento de imágenes e Inteligencia Artificial (específicamente el área de Machine Learning). En el caso del primero, la teoría del HOG se basa en obtener un vector de características que describa la textura de una imagen mediante la distribución de la intensidad de sus gradientes. Para ello, calcula el centrado horizontal y vertical de los gradientes. Luego encuentra las magnitudes y orientación de cada gradiente utilizando las siguientes formulas:

$$g = \sqrt{g_x^2 + g_y^2}$$

$$\theta = \arctan\left(\frac{g_y}{g_x}\right)$$

Después de eso, divide la imagen en pequeñas regiones conectadas llamadas celdas y calcula un histograma de gradientes para cada celda. También para tener en cuenta los cambios en la iluminación y el contraste, los gradientes se normalizan agrupando las celdas en bloques conectados. Finalmente, el descriptor de HOG es el vector concatenado de los histogramas de células normalizadas de todas las regiones de bloque. Entender la teoría del HOG permitió que se puedan ajustar mejor sus parámetros, con el fin de obtener un vector de características más robusto.

Para el segundo, se aplicó como métrica de evaluación el accuracy, con el fin de tener un valor cuantitativo que permita medir la eficiencia del modelo de clasificación obtenido frente a otros modelos de clasificación. Esta métrica se basa en la siguiente formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Donde TP viene a ser los True Positives, TN los True Negatives, FP los False Positives y FN los False Negatives. Además, también se investigo acerca de la teoría de cada uno de los

4 algoritmos de clasificación que se utilizaron dentro de los experimentos, con el fin de ajustar cada uno de sus parámetros.

3.11 Student Outcome K: Diseño y Construcción de Sistemas Complejos

Para la construcción y el desarrollo del sistema de reconocimiento de gestos propuesto, se dividió el proyecto en 4 etapas:

- Investigación
- Diseño
- Desarrollo
- Pruebas

En la primera etapa, se investigó y analizó diversas técnicas enfocadas al reconocimiento de gestos de mano estáticos en los últimos 5 años. Dentro de esa investigación se concluyó que, la mejor técnica para lograr este tipo de reconocimiento, es utilizar un descriptor de imagen (en este caso el HOG) junto con un algoritmo de clasificación, de los cuales se seleccionaron un total de 4 algoritmos.

En la segunda etapa, se diseñó los pasos de cómo se realizaría la extracción del vector de características de una imagen utilizando el HOG y su respectiva experimentación con los algoritmos de clasificación, además de los componentes que tendría el sistema de reconocimiento de gestos de mano utilizando el modelo de clasificación generado dentro de los experimentos.

En la tercera etapa, se pasó a generar un dataset de 3600 imágenes de los gestos estáticos del abecedario de la lengua de señas peruana. Una vez obtenido el dataset, se desarrolló el código fuente del proyecto, donde se implementó las técnicas de sustracción de fondo y del descriptor de imagen HOG, además de realizar un primer modelo de prueba utilizando el dataset recolectado.

En la cuarta fase, se desarrolló el modelo de clasificación utilizando los 4 algoritmos de clasificación seleccionados previamente. Dentro de toda esta fase, se tuvo un énfasis en probar diversos parámetros para cada uno de los algoritmos de clasificación, así como para los parámetros del HOG.

La generación y aplicación de un modelo de reconocimiento de gestos estáticos del abecedario de la lengua de señas peruana, permitió crear un sistema que detecta los 24 gestos

del abecedario de la lengua de señas peruana en tiempo real, utilizando una cámara de la laptop. Este sistema se puede apreciar en la figura 4

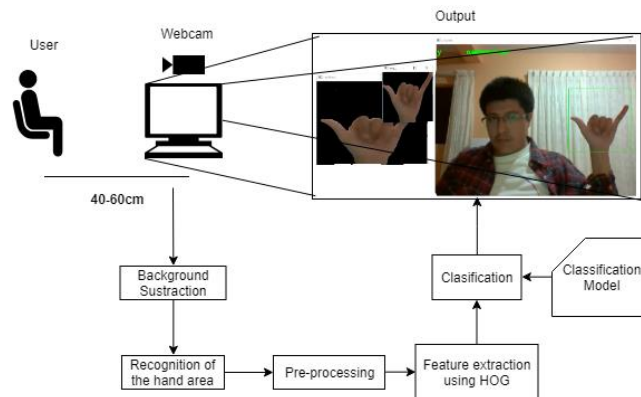


Figura 4: Sistema de reconocimiento de gestos propuesto

Fuente: Elaboración propia

La ventaja de este sistema radica en que permite aprender la lengua de señas peruana de manera interactiva, mediante el deletreo de palabras. Además, este modelo fue construido a partir de las imágenes que se recolectó junto a la asociación ENSEÑAS PERU, lo cual nos permite dar paso a que este dataset pueda ser utilizado para proyectos futuros relacionados a crear juegos que utilicen este tipo de gestos de mano utilizando una tecnología más avanzada (cámaras de profundidad o VR). Adicionalmente, en el ámbito científico, este dataset puede ser aplicado junto a técnicas de Deep Learning, con el fin de analizar los resultados obtenidos con esos algoritmos y determinar si son mejores a los que se obtuvieron con los algoritmos propuestos en este proyecto.

4 ESTADO DE ARTE

El presente capítulo describirá los estudios previos dentro del área de reconocimiento de gestos de mano estáticos en cámaras de baja resolución. El primer tema será sobre segmentación de manos con cámaras de baja resolución, el cual está compuesto por 3 papers. El segundo tema será sobre reconocimiento de gestos de mano mediante descriptores de imágenes, el cual contará con 4 papers. Finalmente, el último tema será sobre algoritmos de clasificación aplicados al reconocimiento de gestos de mano, el cual está conformado por 4 papers.

4.1 Método:

El presente trabajo de investigación surgió debido a la falta de un modelo de clasificación que permita reconocer los gestos de mano estáticos del abecedario de la lengua de señas peruana.

Para plantear la propuesta de esta investigación, primeramente, se identificó los métodos de reconocimiento de gestos de mano existentes, con el fin de encontrar la mejor metodología que se ajuste al problema planteado. A continuación, se utilizó la biblioteca virtual de la universidad para acceder a diversas bases de datos que permitan extraer artículos científicos sobre diversos algoritmos que resuelvan los problemas encontrados previamente. Entre estas bases de datos se encuentran: Scopus, Springer, UTEM y IEEE, siendo esta última la que proveyó una mayor cantidad de artículos.

Se encontró una gran cantidad de artículos sobre el tema, no obstante, solo se consideró aquellos que cumplieran con los siguientes requisitos de investigación: Aporte directo con el tema del proyecto, tiempo de antigüedad de 1 a 4 años y que se encuentren en una fuente confiable. Por consiguiente, se seleccionó los 11 papers más relevantes.

Los 11 artículos seleccionados se dividen en las siguientes tipologías:

- Segmentación de manos con cámaras de baja resolución (3 papers)
- Extracción de características de gestos de mano estáticos mediante descriptores de imágenes (4 papers)
- Algoritmos de clasificación aplicados al reconocimiento de gestos de mano estáticos (4 papers)

Además, en la Tabla 4 se detallan: los títulos de los artículos, nombres de los autores, año de publicación y la fuente donde se consiguió el artículo:

Tabla 4: Recopilatorio de artículos del estado del arte

Tipo	Título	Autor
Segmentación de manos con cámaras de baja resolución	Real time Finger Tracking and Contour Detection for Gesture Recognition using OpenCV	Ruchi Manish Gurav, Premanand K. Kadbe (2015)
	Novel Method for using Hand Recognition as Computer Remote Control with Computer Vision Techniques	Mendoza D., Barreno M., Velasco N. (2017)
	Real-Time Numerical 0-5 Counting Based on Hand-Finger Gestures Recognition	ABD Albary Sulyman, Zeyad T. Sharef, Kamaran Hama Ali Faraj, Zaid Ahmed Aljawaryy, Fahad Layth Malallah (2017)
Extracción de características de gestos de mano estáticos mediante descriptores de imágenes	Hand Gesture Recognition Based on Improved Histograms of Oriented Gradients	Lan Tiantian, Shen Jinyuan, Liu Runjie, Guo Yingying (2015)
	Static hand gesture recognition for Vietnamese sign language (VSL) using principle components analysis	Thao Nguyen Thi Huong, Tien Vu Huu, Thanh Le Xuan, San Vu Van (2015)
	Vision-Based Approach for American Sign Language Recognition Using Edge Orientation Histogram.	Jayshree R. Pansare, Maya Ingle (2016)
	American Sign Language Recognition System Using Image Processing Method	Amit Kumar Gautam (2017)
Algoritmos de clasificación aplicados al reconocimiento de gestos de mano estáticos	Hand Gesture Recognition using Multiclass Support Vector Machine	Md. Hafizur Rahman, Jinia Afrin (2013)
	Hand Gesture Recognition using Deep Convolutional Neural Networks	Gjorgji Strezoski, Dario Stojanovski, Ivica Dimitrovski, Gjorgji Madjarov (2016)
	A novel Approach To American Sign Language Recognition Using Madaline neural Network	Sriparna Saha, Rimita Lahiri, Amit Konar, Atulya K. Nagar (2016)
	Hand gesture recognition using neural network based techniques	Vladislava Bobić, Predrag Tadić, Goran Kvaščev (2016)

Elaboración: Fuente Propia

4.2 Resumen por artículo

A continuación, se presenta un resumen de cada uno de los 11 artículos científicos, considerando en la descripción: La motivación del autor, el aporte del autor en el estudio y el principal resultado:

4.2.1 Segmentación de manos con cámaras de baja resolución

4.2.1.1 Artículo 1

En el artículo “Real time Finger Tracking and Contour Detection for Gesture Recognition using OpenCV” (Gurav M., et. al., 2015), los autores proponen aplicar un método de reconocimiento de gestos de mano, utilizando una cámara de baja resolución, de manera que se pueda reconocer una serie de gestos que permitan la interacción humano-computador.

El principal aporte de los autores fue el de implementar un sistema que permita la detección de 6 gestos estáticos, en base a la cantidad de dedos levantados de la mano, utilizando el algoritmo Adaptive boosting para la detección de manos, el algoritmo de ConvexHull para la detección del contorno de los dedos y el clasificador Haar-like feature, un clasificador interno de Opencv, para entrenar el algoritmo.

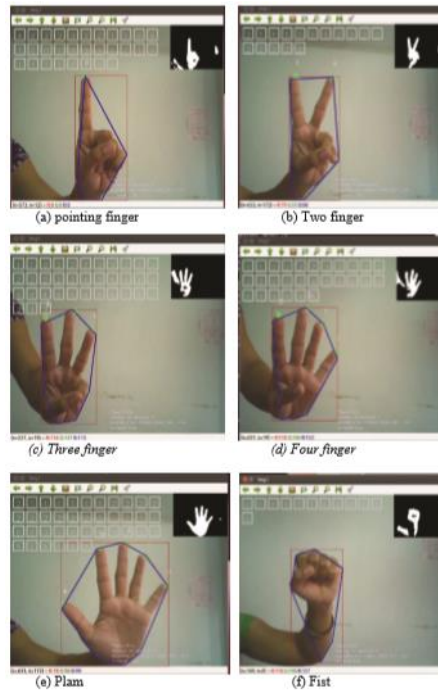


Fig 6: The fingertip & contour detection using Convexhull algorithm

Figura 5: Gestos reconocidos por el sistema

Fuente: Gurav M., et. al., 2015

En base a la experimentación, los autores obtuvieron el siguiente resultado:

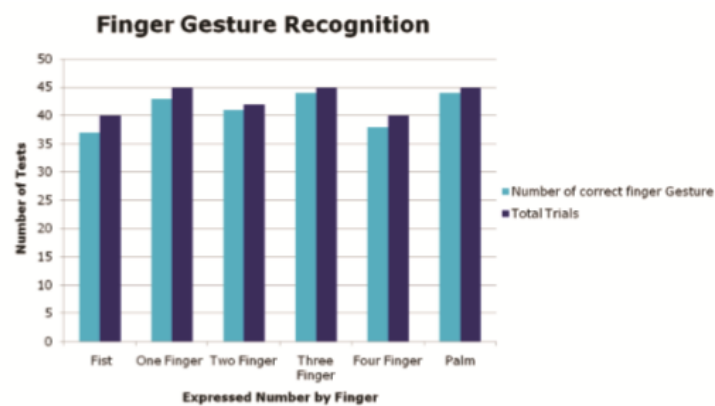


Fig 7: System's performance evaluation results

Figura 6: Resultados del desempeño del sistema

Fuente: Gurav M. et. al., 2015

En base a la figura 6, el sistema obtuvo un accuracy de 92% en el reconocimiento de las manos, lo cual demuestra que el método planteado por los autores permite la detección de gestos de mano con una alta precisión.

Como conclusión, los autores mencionan que: “La principal ventaja de la secuencia de contorno local es que es invariable a la rotación, traslación y escala de la imagen, por lo que es una buena característica para entrenar a la máquina” (Gurav M. ,et. al., 2015).

4.2.1.2 Artículo 2

En el artículo “Novel Method for using Hand Recognition as Computer Remote Control with Computer Vision Techniques” (Mendoza D., et., al., 2017), los autores propusieron mejorar la comunicación de un sistema de interacción humano-computador, utilizando una cámara de baja resolución y gestos estáticos de la mano, con el fin de que se pueda manipular acciones tradicionales de una computadora de manera remota.

El principal aporte de los autores fue el de desarrollar un sistema de reconocimiento de 6 gestos estáticos, de manera que puedan usarse como un control remoto de un computador y así facilitar la interacción humano-computador. Estos gestos estaban conformados por la cantidad de dedos que se estiraban de la mano. De esta manera, los autores lograron registrar un total de 7 acciones distintas, con el fin de utilizarlas en remplazo de un mouse o teclado. Estas acciones se conforman por (Mendoza D., et. al., 2017):

- Mano cerrada (click izquierdo)
- Abrir mano en menos de un segundo (click derecho)
- Mover mano abierta (mover el mouse)
- Un dedo levantado (Enter)
- Dos dedos levantados (Escape)
- Tres dedos levantados (Abrir programas)
- Cuatro dedos levantados (Abrir el reproductor de videos)



Figure 4: Recognition of hand gestures. First row (left hand), Second row (right hand)

Figura 7: Gestos reconocidos por el sistema

Fuente: Mendoza D. et. al., 2017

Como conclusión, los autores mencionan que el algoritmo implementado: “logro que el sistema de reconocimiento permitiera que la detección la región de la mano y los dedos estirados, de manera que no se vea afectado por: la rotación, traslación, posición, color o forma de esta” (Mendoza D., et. al., 2017). Además, el uso de la librería OpenCV logro que el sistema tuviera un alto nivel de respuesta, debido a las funciones optimizadas que incluye esta librería.



Figura 8: Funciones del mouse

Fuente: Mendoza D., et., al., 2017

4.2.1.3 Artículo 3

En el artículo “Real-Time Numerical 0-5 Counting Base don Hand-Finger Gestures Recognition” (Zeyad T., et. al, 2017), los autores desarrollaron un programa que facilite la comunicación de las personas sordas, mediante el reconocimiento de 5 gestos de mano, utilizando una cámara de baja resolución, de manera que pueda ser interpretado por la computadora.

El principal aporte de los autores fue el de implementar una técnica de detección de gestos estáticos y dinámicos de la mano, además del número de dedos levantados, mediante una técnica de escaneo de imágenes en blanco y negro, de manera que pueda detectar los números del 0 al 5 mediante una webcam.

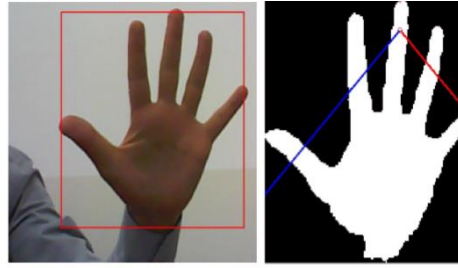


Figure 8: Tracked Hand Gesture Number Five at Left and its Proposed Algorithm at Right.

Figura 9: Reconocimiento de la mano mediante la técnica propuesta

Fuente: Zeyad T., et. al., 2017

En base a la experimentación, el principal resultado de los autores se encuentra en la tabla 5:

Tabla 5: Porcentaje de acierto de la detección de los gestos

Table 1: Overall Results As Match Or Mis-Match Of Hand Gesture Numbers.

Hand Gesture Type	Total Hand Gesture	Total Hand Match	Total Mis-Match	Matching Rate
0	36	36	0	100%
1	30	28	2	93.3%
2	22	22	0	100%
3	12	11	1	91.6%
4	33	33	0	100%
5	47	47	0	100%
Total Gesture Samples	180	177	3	

Fuente: Zeyad T., et. al., 2017

Según esta tabla, los gestos 0,2,4 y 5 obtuvieron un promedio de reconocimiento de 100% mientras que el gesto 3 obtuvo 91.6% y el gesto 1 obtuvo un 93.3%. Según los autores, “esto se debe al ruido que tenía la imagen cuando fue capturada por la cámara”. En promedio, obtuvieron un total de 98.3 % de accuracy.

4.2.2 Extracción de características de gestos de mano estáticos mediante descriptores de imagen

4.2.2.1 Artículo 4

En el artículo “Hand Gesture Recognition Based on Improved Histograms of Oriented Gradients” (Lan Tiantian, et. al., 2015), los autores proponen optimizar el método propuesto por el estudio “An extended HOG model: SCHOG for human hand detection” (X. B. Meng, J. Lin, 2012), de manera que se reduzca el margen de error a la hora de identificar los píxeles de la piel de la mano, dentro de una imagen.

Los principales aportes por parte del autor son:

- Un método para reducir el margen de error a la hora de identificar los píxeles que pertenezcan a la región de la mano, mediante las técnicas de Skin Color Similarity y Histogram of Oriented Gradient (HOG).
- Se logró mejorar el accuracy obtenido en el modelo, mediante la combinación de dos HOG, los cuales contaban con diferentes tamaños de celdas.

El principal resultado obtenido por los autores se encuentra en la tabla 6, los cuales representan el accuracy obtenido en los modelos, comparando los resultados del HOG con la mejora propuesta y sin la mejora propuesta. Para ambos resultados, los autores utilizaron el algoritmo de Support Vector Machine para generar el modelo.

Tabla 6: Resultados de la experimentación

Table1. Recognition rate and time with different hand gestures features

a	Improved HOG		HOG	
	R (%)	T (ms)	R (%)	T (ms)
16	80.17	16.53	70.86	11.80
8	92.93	18.00	89.66	13.65
4	94.48	26.20	91.21	21.38
16+8	92.59	18.81	88.62	14.74
16+4	94.66	26.70	93.10	22.22
8+4	95.34	28.93	93.45	24.37
16+8+4	95.17	29.74	93.28	25.07

Note: a represents the size of cell in HOG is $a \times a$, R represents recognition rate and T represents time of feature extraction.

Fuente: Lan Tiantian, et. al., 2015

La tabla 6 muestra, que el método planteado por los autores mejoró el porcentaje de reconocimiento de los gestos alcanzando hasta una tasa de 95.17 %.

En base a estos resultados, los autores concluyeron que: “Al agregar skin similarity a una magnitud de gradiente se logró mejorar el porcentaje de reconocimiento. Además, las características locales es un factor clave que afecta el reconocimiento del gesto de la mano. Una celda más pequeña es, una característica más fina que describen los histogramas y una mayor tasa de reconocimiento. La combinación de HOG de distintos tamaños puede reflejar características más integrales sobre los gestos de las manos” (Lan Tiantian, et. al., 2015).

4.2.2.2 Artículo 5

En el artículo “Static hand Gesture recognition for Vietnamese Sign Language using principle components analysis” (Thao N., et. al., 2015), los autores desarrollaron un sistema que permita reconocer los gestos del lenguaje de señas de Vietnam, de manera que pueda interpretar los gestos de las personas sordas y así facilitar la comunicación de estas personas con otras que no conocen este lenguaje.

El principal aporte de los autores es un sistema que, utilizando la técnica de Principle Component Analysis (PCA) como extractor de características, se pueda reconocer los gestos del lenguaje de señas de Vietnam, dentro de un fondo estático.



Fig. 2. List of VSL

Figura 10: Lenguaje de señas Vietnamita

Fuente: Thao N., et. al., 2015

En la experimentación, se utilizó 2 algoritmos para clasificar los gestos en base a las características obtenidos por el PCA: La distancia Euclideana y la distancia Mahalanobis. Para clasificar los gestos, se utilizó un dataset de 250 imágenes (10 imágenes para cada gesto). En base a ello se obtuvo el siguiente resultado:

Tabla 7: Porcentaje de reconocimiento del sistema utilizando distancia Euclidiana y Mahalanobis

TABLE II. THE TOTAL SUCCESS RATE OF METHODS USING EUCLIDEAN AND MAHALANOBIS DISTANCE

	Euclidean distance	Mahalanobis distance
Total success rate	90.4%	91.5%

Fuente: Thao N., et. al., 2015

La Tabla 7 muestra que la distancia Mahalanobis presento un mejor accuracy con la técnica propuesta alcanzando un promedio de reconocimiento de 91.5%. En base a estos resultados los autores concluyen que: “Algunas letras recibieron una clasificación incorrecta debido a que existe una gran similitud entre los gestos. No obstante, al utilizar PCA, el sistema logro obtener un alto porcentaje de reconocimiento, y los resultados son mejores cuando se utiliza la distancia Mahalanobis en comparación con la distancia Euclideana.

4.2.2.3 Artículo 6

En el artículo “Vision-Based Approach for American Sign Language Recognition Using Edge Orientation Histogram” (Jayshree R.,et. al., 2016), los autores proponen un sistema que facilite la comunicación del lenguaje de señas americano, mediante un sistema que permita transformar a texto, las letras del abecedario de dicho lenguaje.

El principal aporte de los autores fue el desarrollo de un sistema de reconocimiento de gestos del abecedario del lenguaje de señas americano, con el fin de traducir a texto (de la “A” a la “Z”), los gestos realizados frente a una webcam, mediante la técnica de Edge Orientation Histogram (EOH), con imágenes que se encuentren en un fondo complejo y con distintos tipos de luz.

En base a la experimentación, el principal resultado de los autores se encuentra en la tabla 8:

Tabla 8: Resultados del Sistema A-ASLR

TABLE I. EXPERIMENTAL RESULTS OF A-ASLR SYSTEM

ASL Symbol	Recognition Rate (%)	ASL Symbol	Recognition Rate (%)
A	95	N	80
B	90	O	80
C	90	P	90
D	90	Q	80
E	90	R	80
F	90	S	90
G	90	T	90
H	90	U	90
I	90	V	90
J	90	W	90
K	90	X	80
L	90	Y	90
M	90	Z	90

Fuente: Jayshree R. et. al. 2016

Los autores demostraron que el sistema propuesto obtuvo un promedio de reconocimiento de 88.26% en un tiempo de ejecución de 0.5 segundos para las 26 letras del lenguaje de señas americano. Además, la metodología propuesta logra extraer características de imágenes de baja calidad, lo cual permite que pueda usarse con cámaras de baja resolución.

4.2.2.4 Artículo 7

En el artículo “American Sign Language Recognition System Using Image Processing Method” (Kumar A., 2017), los autores proponen reconocer los 24 gestos estáticos del abecedario del lenguaje de señas americano, mediante una nueva técnica basado en la detección de puntos de referencia de la mano.

El principal aporte de los autores fue una técnica de reconocimiento de gestos del lenguaje de señas americano, la cual se basa en la detección del centroide de la mano, de manera que, utilizando como clasificador la distancia euclidiana, se pueda calcular la distancia hacia los bordes de la mano y generar un descriptor que cubra toda la región de la mano.

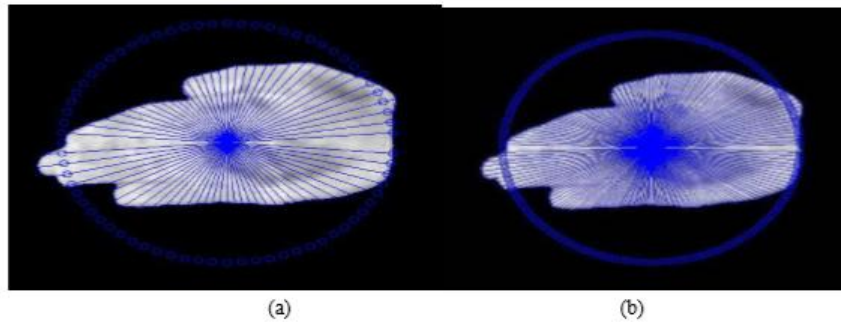


Figure 5. Landmark distance from centroid at (a) 5 Degree (b) 2 Degree.

Figura 11; Distancia entre los centroides y el borde de la mano

Fuente Kumar A. et. al, 2017

El principal resultado está compuesto por 2 descriptores, los cuales están conformados por una serie de rectas que representan la distancia entre el centroide y el borde de la mano. Para este caso, los 2 descriptores que obtuvieron el mejor resultado se encuentran en la tabla 9.

Tabla 9: Accuracy obtenido

TABLE 2: SUMMARIZED ACCURACY RATE

Method	Images in Training Set	Correctly recognized	Recognition Rate
72 point Descriptor	164	123	75.0%
180 point Descriptor	164	131	79.9%

Fuente Kumar A. et. al., 2017

En base a estos resultados, los autores concluyen que: “Al aumentar la cantidad de puntos de descriptores de 72 a 180, se mejoró el porcentaje de reconocimiento. No obstante, a pesar de que el porcentaje de reconocimiento es relativamente bajo (79.9%), el algoritmo implementado tarda menos de medio segundo en realizar el reconocimiento. Además, el algoritmo no se ve afectado por la rotación, luminosidad o traslación de la mano y no necesita de un equipo externo (guante) para que funcione (Kumar A. et. al., 2017).

4.2.3 Algoritmos de clasificación aplicados al reconocimiento de gestos de mano estáticos

4.2.3.1 Artículo 8

En el artículo “Hand Gesture Recognition using Multiclass Support Vector Machine” (Rahman H., Jinia Afrin, 2013), los autores desarrollaron un sistema de reconocimiento de

gestos de mano, de manera que pueda facilitar la interacción humano- computador, con el fin de que estos gestos puedan transmitir un tipo de acción en el programa.

El principal aporte de los autores es: Un sistema de reconocimiento de gestos de mano, que permita reconocer 10 gestos conformados por letras del abecedario del lenguaje de señas americano, utilizando la técnica de Biorthogonal Wavelet Transformation y el algoritmo de clasificación Multi-Class Support Vector Machine.



Fig 4: Finger spelled Alphabet (Top row [A, B, C, D, G], Bottom row [H, I, L, V, Y]).

Figura 12: Gestos a reconocer por el sistema

Fuente: Rahman H., Jinia Afrin, 2013

En base a la experimentación, el principal resultado fue la siguiente tabla, la cual se basó en un modelo entrenado por 2300 imágenes utilizando la metodología propuesta por los autores.

Tabla 10: Resultados del reconocimiento de 10 gestos para 5 usuarios

Table 1. Recognition results of 10 gestures of 5 users

User \ Gesture	A	B	C	D	G	H	I	L	V	Y
1	93%	96%	88%	83%	94%	82%	88%	92%	88%	73%
2	92%	90%	85%	88%	92%	90%	85%	90%	82%	82%
3	88%	92%	82%	90%	87%	90%	92%	90%	93%	87%
4	86%	94%	84%	92%	90%	88%	82%	91%	90%	82%
5	75%	88%	88%	93%	90%	82%	80%	88%	84%	80%
Total	87%	92%	85%	90%	91%	86%	85%	90%	87%	81%

Fuente: Rahman H., Jinia Afrin, 2013

En base a estos resultados, los autores demostraron que el sistema propuesto logro detectar los 10 gestos, en base a la experimentación con 5 personas, con un promedio de accuracy de 87.4%. En base a ello, los autores concluyeron que “existen diversos tipos de algoritmos de clasificación como redes neuronales y árboles de clasificación. No obstante, el algoritmo de

Support Vector Machine presenta ser un buen candidato para la clasificación de gestos de mano. Esto se debe a su alto rendimiento de generalización, sin la necesidad de agregar un conocimiento previo, incluso cuando la dimensión del espacio de entrada es muy alta.” (Rahman H., Jinia Afrin, 2013).

4.2.3.2 Artículo 9

En el artículo “Hand Gesture Recognition using Deep Convolutional Neural Networks” (Strezoski G., et. al., 2016), los autores compararon 4 arquitecturas de redes convolucionales de estudios previos, dentro del área de reconocimiento de manos, además de proponer una arquitectura propia, de manera que presente un modelo robusto y eficiente.

Los principales aportes de los autores fueron (Strezoski G., et. al., 2016):

- “Evaluamos varias redes neuronales convolucionales simples y pre entrenadas en diferentes conjuntos de datos y comparamos su desempeño”.
- “Formamos un modelo robusto y profundo para reconocimiento de gestos con una alta precisión”.
- “Reportamos un buen desempeño de forma temporal con solo 2 ms de tiempo de clasificación en un modelo completamente entrenado, lo que lo convierte en un modelo funcional en tiempo real”.

El principal resultado se basó en 2 tablas:

La primera tabla corresponde al accuracy obtenido para cada una de las arquitecturas investigadas en este estudio, además de la arquitectura propuesta por los autores. Las cuales, corresponden a un modelo que permite reconocer 10 gestos de la mano.

Tabla 11: Accuracy de las arquitecturas

Table 3: Accuracy scores for each models best run

	Top-1	Top-3
GoogLeNet	78.22%	90.41%
AlexNet	42.18%	60.9%
Custom model	64.17%	84.32%
LeNet	28%	47.19%
VGG model	64.19%	83.33%

Fuente Strezoski G., et. al., 2016

La segunda tabla corresponde al tiempo de entrenamiento, para cada una de las arquitecturas investigadas en este estudio. En ella se consideró el tiempo en minutos para cada epoch y el tiempo en ms para clasificar una imagen.

Tabla 12: Duración del epoch de las arquitecturas de CNN (con GPU)

Table 4: Average classification and training epoch duration on a GPU

	Training epoch (min)	Classification per image (ms)
GoogLeNet	4	2.8
AlexNet	2	1
Custom model	2.5	0.5
LeNet	0.4	0.6
VGG model	5	2

Fuente Strezoski G., et. al., 2016

En base a estos resultados los autores concluyeron que: “El modelo propuesto no obtuvo la mejor precisión en comparación a la arquitectura GoogLeNet. No obstante, fue el que logro una mayor rapidez al procesar una imagen, lo cual hace que esta arquitectura sea más eficiente en sistema de control, como la interacción humano-robot, donde la velocidad de respuesta es importante” (Strezoski G., et. al., 2016).

4.2.3.3 Artículo 10

En el estudio “A novel Approach To American Sign Language Recognition Using Madaline neural Network” (Saha S., et. al., 2016), los autores proponen una nueva arquitectura de red neuronal, de manera que esta permita identificar los gestos del abecedario del lenguaje de señas americano.

Los principales aportes de los autores son:

- Una nueva arquitectura de red neuronal (MAdaline), basada en la red Adaline para la clasificación de gestos de mano.
- Una técnica de extracción de características, basada en el estudio de “A leaf recognition algorithm for plant classification using probabilistic neural network” (Wu S. G. et. al , 2007), aplicado para el reconocimiento de gestos de mano.

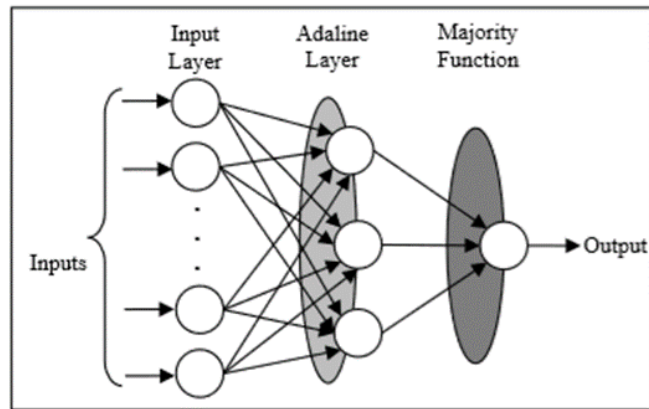


Fig. 2. Architecture of MAdaline neural network.

Figura 13: Arquitectura de la red MAdaline

Fuente Saha S. et. al., 2016

El principal resultado se basó en la comparación de las variables: Precision, Recall, Accuracy y F1Score contra otros algoritmos de clasificación.

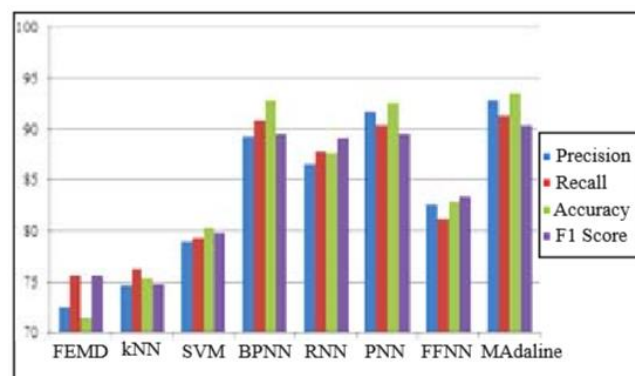


Fig. 7. Comparison of proposed work with existing literatures.

Figura 14: Comparación de MAdaline con otros algoritmos

Fuente Saha S. et. al., 2016

En base a estos resultados, se demostró que el algoritmo propuesto logró obtener un mejor resultado en comparación a otros algoritmos de clasificación, lo cual demuestra que la red MAdaline, junto a las características presentadas por los autores, permite obtener un modelo que sea robusto y de alta precisión.

Como conclusión, los autores mencionan que: “El algoritmo propuesto no requiere pasos adicionales para extraer información relevante del background. No obstante, aunque las características derivadas aquí requieren conceptos geométricos elementales, estas

características son más sofisticadas y fáciles de tratar que las características geométricas popularmente usadas. Además, el cálculo de tales características implica el procesamiento de múltiples características sin procesar y como consecuencia estas características son más ricas en términos de contenido de información”. (Saha S. et. al. 2016).

4.2.3.4 Artículo 11

En el artículo “Hand gesture recognition using neural network based techniques” (Bobic V., et. al., 2016), los autores proponen aplicar 2 arquitecturas de redes neuronales, con el fin de clasificar 10 gestos de mano estáticos, de manera que el modelo generado pueda reconocer estos gestos en fondos estáticos y con imágenes de distinta orientación.

Los principales aportes de los autores se basaron en 2 arquitecturas de redes neuronales, utilizando como datos de entrada un vector de características obtenidas por un Histogram Oriented Gradient, de manera que generen un modelo de clasificación para reconocer 10 gestos de mano estáticos. Estas arquitecturas fueron:

- Un MultiLayer Perceptron conformado por 1 hidden layer de 15 neuronas
- Un MultiLayer Perceptron, conformado por 5 hidden layers, donde los datos de entrada son codificados de manera que al atravesar cada layer de la red neuronal se aplica una técnica de “Softmax Layer” de manera que se decodifique los datos y se puede identificar el gesto.

En base a la experimentación, los autores obtuvieron el siguiente resultado para la 1° arquitectura.

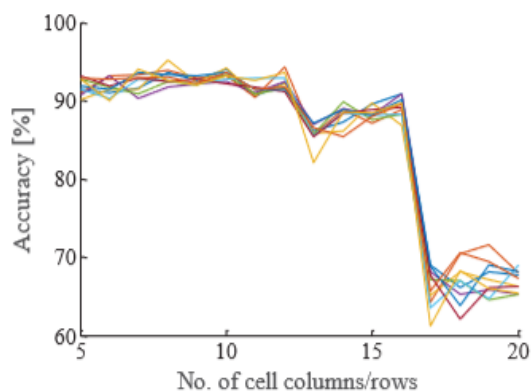


Fig. 6. Accuracy of performed classification depending on the cell size, for 10 repetitions.

Figura 15: Accuracy del MLP en base al n° de celdas usadas en el HOG

Fuente: Bobic V. et. al., 2016

Y, para la 2° arquitectura los autores obtuvieron la siguiente tabla como resultado:

TABLE I: DESCRIPTIVE STATISTICS FOR ACCURACY OF PERFORMED SPARSE AUTOENCODER BASED CLASSIFICATION FOR DIFFERENT LENGTHS OF FEATURE SET, AVERAGED OVER 10 REPETITIONS

<i>No. of layers</i>	<i>No. of parameters</i>	<i>Classification Accuracy [%]</i>
1	800	88.6 ± 0.4
2	400	92.5 ± 0.2
3	200	90.9 ± 1.0
4	100	88.6 ± 0.4
5	50	87.9 ± 1.2

Figura 16: Accuracy del Sparse Autoencoder en base al n° de parámetros

Fuente: Bobic V. et. al., 2016

En ambos casos, los métodos aplicados por los autores obtuvieron una clasificación mayor a 90%, lo cual demuestra que utilizar un HOG descriptor permite el reconocimiento de gestos de la mano sin que estas se vean afectadas por la luz, fondo o tono de piel. Por otro lado, “utilizar un autoencoder permitió obtener características de diferentes niveles de abstracción permitiendo reducir la dimensionalidad de las imágenes” (Bobic V., et. al., 2016).

4.3 Conclusiones

4.3.1 Conclusiones para la tipología de segmentación de manos con cámaras de baja resolución

Se identificó que los papers N° 1, N° 2 presentan una técnica en común para localizar la ubicación de la mano... “Convex Hull”. Esta técnica, permite encerrar la región de la mano, de manera que el algoritmo de reconocimiento pueda aplicarse dentro de la región encerrada. Por otro lado, el paper N° 3 presenta un enfoque distinto al escanear una imagen en blanco y negro, lo cual hace que presente un mejor accuracy a coste de un mayor tiempo de procesamiento. En ambos casos, estas técnicas lograron reconocer el gesto que realizaba la mano sin la necesidad de algún equipo externo (guante) o una cámara costosa (kinect). No obstante, la cantidad de gestos reconocidos estuvo limitada a la cantidad de dedos que

levantaba la mano, lo cual demuestra que, a pesar de ser buenas técnicas para reconocer gestos de mano estáticos, no son suficientes para abarcar una mayor cantidad de gestos.

4.3.2 Conclusiones para la tipología de extracción de características de gestos de mano estáticos mediante descriptores de imágenes

Se identificó que los papers N°4 y N°5 presentaron descriptores enfocados en técnicas de detección de orientación de gradiente en una imagen. Dentro de estos papers se puede observar que el paper N° 4 obtiene un mejor accuracy utilizando la técnica de Histogram Oriented Gradient (HOG) en contraste con el paper N° 5 el cual utiliza Edge Oriented Gradient (EOG). Esto se debe principalmente, al aporte presentado por el paper N° 4 donde su principal resultado muestra una mejora significativa en el accuracy obtenido utilizando el nuevo HOG.

Para el paper N° 6 se puede observar que la técnica propuesta logro identificar todos los gestos de manera rápida, además de clasificar el gesto independientemente de la orientación que este tenía. No obstante, el bajo accuracy que presenta (79.9%) solo lo hace eficiente en sistemas donde se priorice la velocidad de respuesta en vez de la precisión.

Para el paper N° 7 se puede observar que la técnica propuesta por los autores logro identificar una mayor cantidad de gestos utilizando PCA, además de obtener un buen accuracy. No obstante, la baja cantidad de instancias utilizadas por gesto (10 imágenes) genera el riesgo de que el modelo presente un *overfitting*.

4.3.3 Conclusiones para la tipología de algoritmos de clasificación aplicados al reconocimiento de gestos de mano estáticos

En esta tipología se pudo identificar que los algoritmos de clasificación aplicados dentro del área de reconocimiento de gestos de manos estáticos son: las redes neuronales (papers N°9, N°10 y N°11) y el algoritmo de Support Vector Machine (SVM) (paper N°8). También se identificó que los datos de entrada de estos algoritmos corresponden a la matriz de la imagen o un descriptor de características (HOG en el paper N° 11). Finalmente, los 4 papers presentados en esta tipología lograron obtener un modelo que presenta un alto accuracy para el reconocimiento de gestos de mano estáticos, lo cual demuestra que estos algoritmos son buenas alternativas para entrenar un modelo de clasificación.

5 MARCO TEÓRICO

En este capítulo se definirá los conceptos de reconocimiento de gestos de mano, lo cual involucra el preprocesamiento y la metodología que se usará (descriptor de imagen). Además, se describirá los conceptos relacionados a *Machine Learning*, además del algoritmo de clasificación que se piensa utilizar.

5.1 Reconocimiento de gestos de mano

“El reconocimiento de gestos de mano corresponde al uso de la mano y/o brazo para producir una serie de posturas o movimientos, (...) con el fin de ofrecer una interacción simple y eficiente para la comunicación Humano-Computador” (Bei Li, et. al., 2017).

Existen 2 tipos de reconocimiento de gestos (Bei Li, et. al., 2017):

- Reconocimiento de gestos de mano estáticos: Los cuales corresponden a las características de la forma de los gestos de la mano en un tiempo determinado.
- Reconocimiento de gestos de mano dinámicos: Representan una serie de acciones de la mano en un determinado periodo de tiempo, lo cual aumenta la información del tiempo y las características del movimiento.

5.2 Reconocimiento basado en visión computacional

El reconocimiento de gestos basado en visión computacional utiliza una o más cámaras para recolectar imágenes o vides de gestos, con el fin de utilizar métodos de procesamiento de imágenes y *Machine Learning* para analizar las características de estos gestos. De este método es que el dispositivo de entrada es barato, y la cámara es cada vez más popular en todo tipo de productos electrónicos, y no requiere requisitos adicionales para la mano humana. La interacción entre la computadora y el ser humano es más natural. Por lo tanto, la investigación sobre el reconocimiento de gestos de la mano basado en la visión computacional es cada vez más; lo que tiene una gran mejora en la tasa de reconocimiento y el rendimiento en tiempo real. (Bei Li, et. al., 2017).

5.3 Procesamiento de imágenes

El procesamiento de imágenes utiliza una serie de operaciones matemáticas para cualquier forma de procesamiento de señales, la cual la entrada es: una imagen, una secuencia de

imágenes o un video. La salida del procesamiento viene a ser una imagen o un conjunto de características o parámetros relacionados a ella (Gonzalez C. et. al., 2008). Debido a que la imagen de una mano puede presentar: distintos tonos de piel, ruido, luminosidad y presentar un fondo no uniforme, es necesario que esta reciba un tipo de preprocesamiento que permita facilitar la extracción de las características más importantes. Las técnicas de preprocesamiento permiten aplicar diversos filtros sobre una imagen, de manera que se pueda obtener un cierto grado de características que faciliten la identificación del gesto.

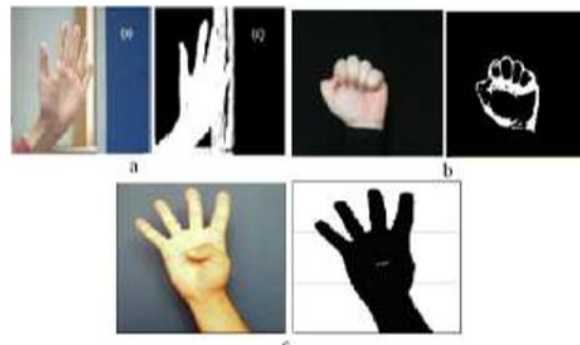


Figura 17: Filtros de pre-procesamiento de mano

Fuente: Pareek A. et. al, 2013

5.4 Descriptores de imagen

Una técnica de extracción de características es el descriptor de imagen, el cual permite conseguir una serie de puntos de interés dentro de una imagen, con el fin de obtener un vector de características que explique el comportamiento de los píxeles dentro de la imagen.

Una vez que se ha detectado un conjunto de puntos de interés a partir de una imagen en una ubicación $p(x, y)$, escala (s) y orientación (θ), su contenido o arquitectura simétrica se necesita codificar en un descriptor adecuado para la comparación discriminativa e insensible a las deformaciones de la imagen local. El descriptor debe estar alineado con θ y proporcional a la escala(s) (Hassaballah M, et. al., 2016).

Las ventajas de utilizar estos descriptores son: (Hassaballah M, et. al., 2016)

- **Robustez:** El algoritmo de detección de características es capaz de detectar las mismas características independientes de escala, rotación, desplazamiento, deformaciones fotométricas, artefactos de compresión y ruido.

- Repetibilidad: El algoritmo de detección de características debe ser capaz de detectar repetidamente las mismas características de la misma escena u objeto bajo una variedad de condiciones de visualización.
- Precisión: El algoritmo de detección de características debe localizar con precisión las características de la imagen (las mismas ubicaciones de píxeles), especialmente para tareas de coincidencia de imágenes, donde se necesitan correspondencias precisas para estimar la geometría epipolar.
- Eficiencia: El algoritmo de detección de características debería ser capaz de detectar características en nuevas imágenes rápidamente para admitir aplicaciones en tiempo real.
- Cantidad: El algoritmo de detección de características debe ser capaz de detectar todas o la mayoría de las características de la imagen. Donde, la densidad de las características detectadas debe reflejar el contenido de información de la imagen para proporcionar una representación compacta de la imagen.

5.5 Histogram Oriented Gradient

Un Histogram Oriented Gradient (HOG) es un descriptor de características, que permite detectar objetos en base a la textura de la imagen. La técnica del descriptor HOG cuenta la cantidad de ocurrencias de la orientación de la gradiente, de cierto keypoints (píxeles), localizados dentro de una región de interés.

La implementación del HOG descriptor es el siguiente (Intel, 2017)

- Dividir la imagen en pequeñas regiones conectadas (celdas) y, para cada celda, calcular el histograma de direcciones de gradiente para los píxeles dentro de la celda.
- Discretizar cada celda en contenedores angulares de acuerdo a la orientación del gradiente.
- El píxel de cada celda contribuye al gradiente ponderado a su contenedor angular correspondiente.
- Los grupos de celdas adyacentes se consideran regiones espaciales llamadas bloques. La agrupación de celdas en un bloque es la base para agrupar y normalizar los histogramas.
- El grupo de histogramas normalizados representa el histograma bloque. Por ende, el conjunto de estos histogramas bloque representa el descriptor.

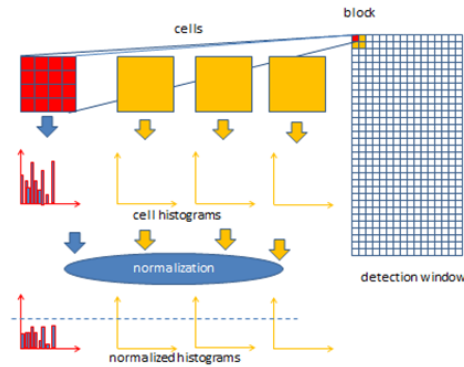


Figura 18: Descriptor HOG

Fuente: Intel 2017

5.6 Lengua de señas

El lenguaje de señas, es una lengua creada por la comunidad de sordomudos, con el fin de facilitar la comunicación de las personas con discapacidad de escuchar u oír. No obstante, este lenguaje no es universal, por lo que algunos gestos pueden tener un significado distinto en determinado país o región.

Para articular las letras del abecedario, se utiliza una serie de gestos realizados por las manos, los cuales representan una letra en específico. En caso se quiera articular una palabra, estos gestos vienen acompañado de movimientos con los brazos y/o expresiones faciales.



Figura 19: Abecedario del lenguaje de señas peruano

Fuente Ministerio de Educación del Perú 2010

5.7 Machine Learning

Machine Learning es una rama de la inteligencia artificial, el cual tiene como finalidad el desarrollar técnicas de aprendizaje para un computador. De esta manera, se busca generar un modelo algorítmico, a partir de un conjunto de información, que permita generalizar el comportamiento de una máquina.

Para el reconocimiento de manos, el tipo de aprendizaje que más se utiliza es el aprendizaje supervisado, el cual en base a una muestra de entrada-salida, llamado la muestra de entrenamiento, la tarea es encontrar una función determinista, que mapee cualquier entrada a una salida, con el fin de predecir futuras observaciones, minimizando el error tanto como sea posible. (Paulo Trigueiros et. al., 2016).

5.8 Support Vector Machine

Support Vector Machine (SVM) es un algoritmo de clasificación que traza cada elemento de los datos como un punto en el espacio n-dimensional (donde n es el número de características que tiene) y el valor de cada característica es el valor de una determinada coordenada. Luego, se realiza la clasificación al encontrar el hiperplano que permite diferenciar entre 2 clases (Sunil Ray, 2017).

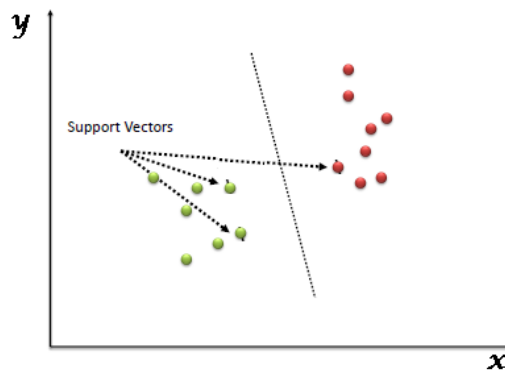


Figura 20: Plano de separación

Fuente Sunil Ray 2017

6 DESARROLLO DEL PROYECTO

El presente capítulo describirá la metodología utilizada dentro del proyecto, la cual incluye:

- La arquitectura del sistema propuesto y las fases que lo componen
- Las técnicas y algoritmos utilizados para la sustracción de fondo, seguimiento y reconocimiento de los gestos de mano
- La metodología aplicada para armar el dataset

6.1 Sistema propuesto

El sistema permite reconocer 24 gestos estáticos del abecedario de la lengua de señas peruana. Este sistema se divide en 5 fases, los cuales se pueden observar en la Figura 21.

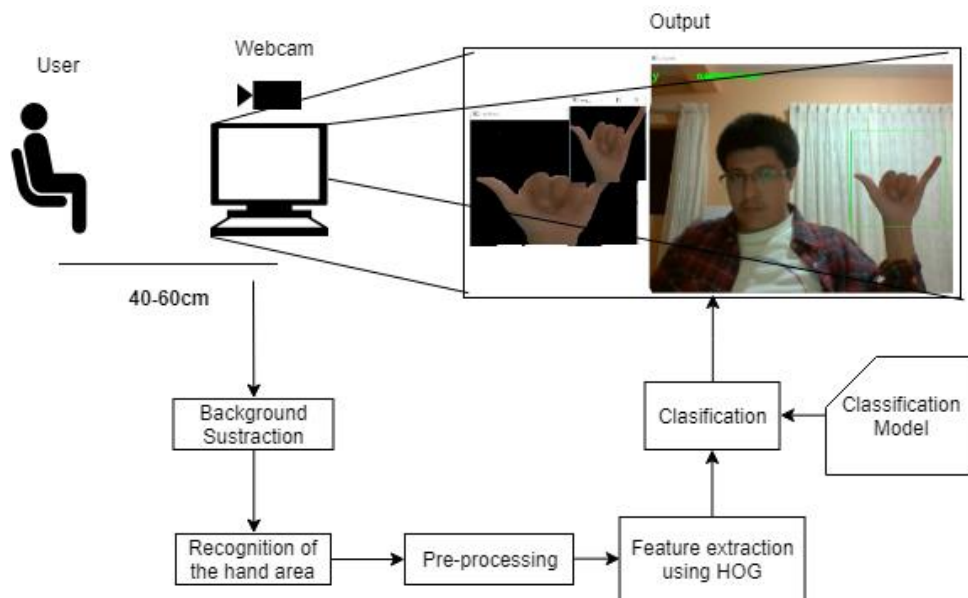


Figura 21: Fases del sistema de reconocimiento de gestos propuesto

Fuente: Elaboración propia

En la primera fase, se aplicó el algoritmo de “Gaussian Mixture-based Background Segmentation Algorithm” (Zoran Z., et. al., 2006), utilizando OpenCV, de manera que se pueda aislar el fondo, de la imagen y se facilite el reconocimiento de la mano. Este algoritmo se ejecutó dentro de un entorno iluminado y con un fondo estático, con el fin de minimizar la mayor cantidad de ruido.

La segunda fase corresponde al reconocimiento del área de la mano. Para ello, el usuario se coloca entre unos 40 a 60 centímetros de la cámara de la laptop y levanta su mano dentro del recuadro verde como se aprecia en la Figura 21. Una vez levantado la mano, el programa

detecta el área de la mano, utilizando el algoritmo de convexHull, y procede a recortar la región de interés y redimensionar la imagen.

La tercera fase corresponde a la etapa de preprocesamiento de la región de interés, el cual consiste en: Aplicar un filtro bilateral, con el fin de eliminar el ruido de la mano; llevar la imagen a escala de grises y aplicarle un threshold con el fin de uniformizar los pixeles negros.

La cuarta fase aplica un descriptor global, debido a que el único elemento presente en la imagen es la región de la mano. Para ello, se utiliza el algoritmo de Histogram Oriented Gradient (HOG) (Dalal N. et. al, 2005) utilizando OpenCV. En este algoritmo la distribución (histograma) de la dirección de las gradientes (gradientes orientadas) es usada como un vector de características. Primero, el algoritmo calcula las gradientes horizontales y verticales. Luego encuentra las magnitudes y la orientación de la gradiente usando las siguientes formulas:

$$g = \sqrt{g_x^2 + g_y^2}$$

$$\theta = \arctan\left(\frac{g_y}{g_x}\right)$$

Donde g es la magnitud de la gradiente y θ es la orientación de la gradiente. Después de eso, divide la imagen en (i x i) celdas y calcula el histograma de gradientes para cada celda. Dependiendo del número bins, el histograma se extiende en 0°- (20° x bins). Además, para considerar los cambios en la iluminación y el contraste, las intensidades de las gradientes deben estar normalizadas, para lo cual se requiere agrupar las celdas en bloques espacialmente conectados. Finalmente, el descriptor HOG es el vector concatenado de los componentes normalizados de los histogramas de las celdas de todas las regiones de los bloques.

La quinta fase corresponde a la clasificación del gesto utilizando el vector de características previamente extraído y el modelo de clasificación entrenado. El gesto clasificado se muestra en la esquina superior izquierda de la pantalla.

6.2 Generación del dataset

La metodología aplicada para la construcción de este dataset se puede visualizar en la Figura 22:

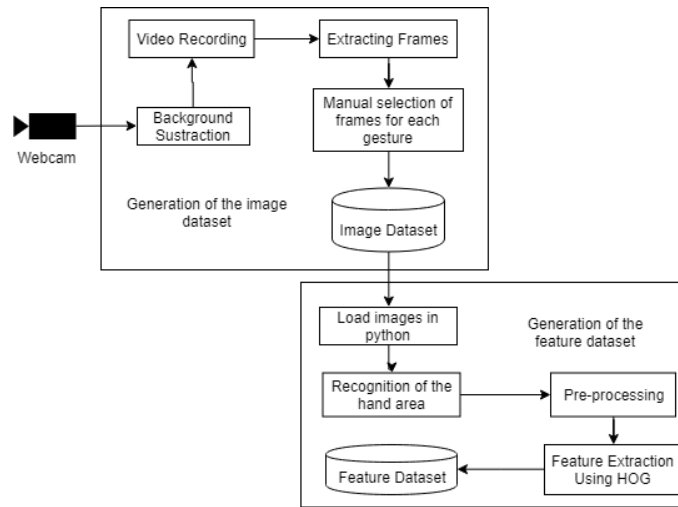


Figura 22: Metodología para generar el dataset

Fuente: Elaboración propia

La construcción del dataset se dividió en 2 etapas:

La primera consistió en la generación del dataset de imágenes de los 24 gestos del abecedario de la lengua de señas peruana. Con el fin de crear un dataset robusto, se grabó 15 videos de 15 personas realizando los 24 gestos estáticos, con el fin de extraer los frames donde se pueda apreciar cambios de rotación y escala en la mano. Para lograr esto, primero se extrajo el fondo de la imagen utilizando el algoritmo de “Gaussian Mixture-based Background Segmentation Algorithm” (Zoran Z., et. al., 2006). Después, se realizó la grabación de los 24 gestos estáticos del abecedario de la lengua de señas peruana. Para ello, se pidió a los participantes que levantaran su mano dentro del recuadro verde y realizaran pequeños movimientos, entre los cuales se incluye: rotación en el eje “X” y “Y” y alejar y acercar su mano a la cámara. Finalmente, una vez terminada las grabaciones, se realizó la partición del video en frames y se seleccionó manualmente las imágenes para cada gesto. En esta selección se consideró las imágenes donde se aprecie cambios de rotación y escala.

La segunda etapa consistió en la generación del dataset de características, el cual nos permitió entrenar los algoritmos de clasificación. Para ello, una vez generado el dataset de imágenes, se procedió a cargar cada una de las imágenes del dataset dentro de Python y se

reconoció el área de la mano utilizando el algoritmo de convexHull. Una vez detectada esta región, se recortó y redimensiono las imágenes a un tamaño de 200x200 píxeles, con el fin de mantener un formato uniforme dentro del dataset. La Figura 23 muestra la mano antes y después del proceso de recorte y redimensión.

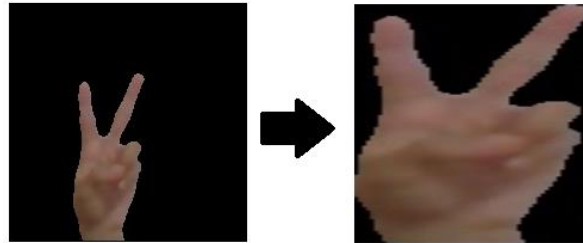


Figura 23: Detección, recorte y redimensión de la región de la mano

Fuente Elaboración propia

Finalmente, se aplicó un filtro gris a las imágenes y se generó el vector de características utilizando el HOG.

6.3 Hardware y Software

Para la experimentación se utilizó una webcam integrada de 720p y una resolución de 4:3, la cual también se utilizó para la generación del dataset. El sistema se probó en una laptop Intel Core i7-7770HQ CPU con 2.80Ghz y 16GB de RAM. El OS fue Windows 10 de 64-bits. El sistema de reconocimiento fue desarrollado en Python 3.5 dentro del IDE Visual Studio Code. Para el desarrollo de los algoritmos de preprocesamiento, sustracción de fondo y HOG, se utilizó la librería OpenCV. Además, para facilitar el proceso de análisis de datos y la generación de modelos utilizando algoritmos de clasificación, se utilizó la librería scikit-learn (Pedregosa et al., 2011).

6.4 Dataset de Imágenes

El dataset fue generado con el apoyo de 15 personas sordas, las cuales se les grabó realizando cada uno de los 24 gestos estáticos del abecedario de la lengua de señas peruana. Las grabaciones se realizaron en un ambiente iluminado y en un fondo estático. Estas grabaciones registran gestos estáticos, considerando pequeñas variaciones de rotación en el eje “X” y en el eje “Y”, con el fin de obtener un dataset robusto a estos cambios. Una vez realizado la grabación, se extrajo todos los frames del video y se seleccionó un total de 3600

imágenes, donde estuvieran presentes las variaciones mencionadas. Cada uno de los 24 gestos contiene 150 imágenes (15 por persona).

6.5 Dataset de Características

Para la generación del dataset de características las imágenes fueron redimensionadas a un tamaño de 64x48, con el fin de obtener un descriptor global utilizando el HOG. La configuración del HOG se puede observar en la Tabla 13.

Tabla 13: Parámetros del HOG

Parámetro	Valor
Tamaño de ventana	64x48
Tamaño del bloque	8x8
Separación de bloque	4x4
Tamaño de la celda	8x8

Fuente: Elaboración propia

6.6 Entrenamiento y Validación

Para el entrenamiento de los modelos se utilizó 4 algoritmos provistos por la librería scikit-learn, los cuales fueron:

- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Random Forest (RF)
- MultiLayer Perceptron (MLP)

Para la validación de estos algoritmos, se utilizó *15-fold cross-validation*, teniendo como métrica de evaluación el accuracy, debido a que se está utilizando una misma cantidad de instancias para cada gesto (150). Para obtener el valor del accuracy se utilizó la siguiente formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

7 RESULTADOS DEL PROYECTO

El presente capítulo describirá los resultados obtenidos en la experimentación para generar el modelo de clasificación:

Los resultados de la experimentación se basaron en los parámetros presentados en el capítulo 6 de la memoria, al cual se le altero el número de bins, con el fin de saber si un descriptor con más o menos características proporcionaba un mejor accuracy. A continuación, las siguientes imágenes muestran los resultados obtenidos para cada uno de los algoritmos en base a la cantidad de bins utilizados.

7.1.1 Support Vector Machine

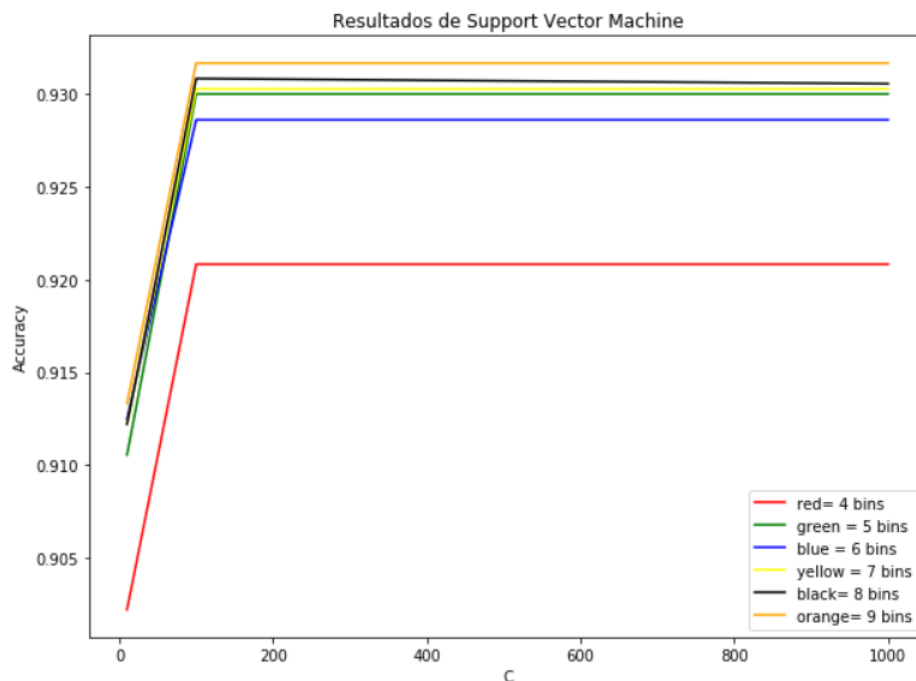


Figura 24: Resultados de Support Vector Machine utilizando HOGs de 4 a 9 bins

Fuente: Elaboración propia

En la Figura 24 se puede observar que el modelo obtiene un mejor accuracy a medida que el parámetro de C se acerca a 100, para después seguir un accuracy. Además, se puede observar que al aumentar de bins, se logró obtener un mejor accuracy en el modelo.

El mejor modelo de este algoritmo obtuvo un accuracy de 93.16%, el cual se logró utilizando un HOG de 9 bins y un C de 1000. En la Figura 25 se puede observar la matriz de confusión de este modelo, donde se observa que las letras con mayor índice de confusión fueron: “k”, “m”, “n”, “o”, “p”.

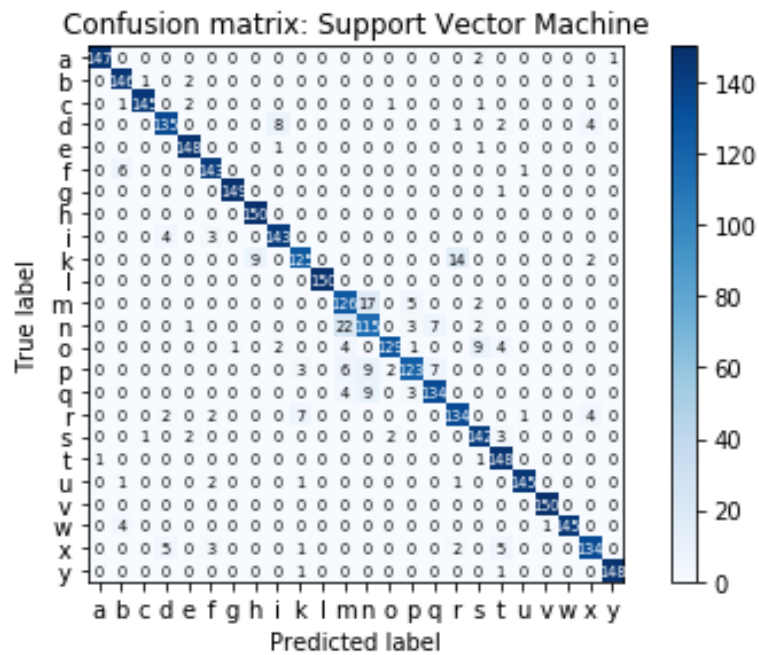


Figura 25: Matriz de confusi3n de SVM con un HOG de 9 bins

Fuente: Elaboraci3n propia

7.1.2 K-Nearest Neighbors

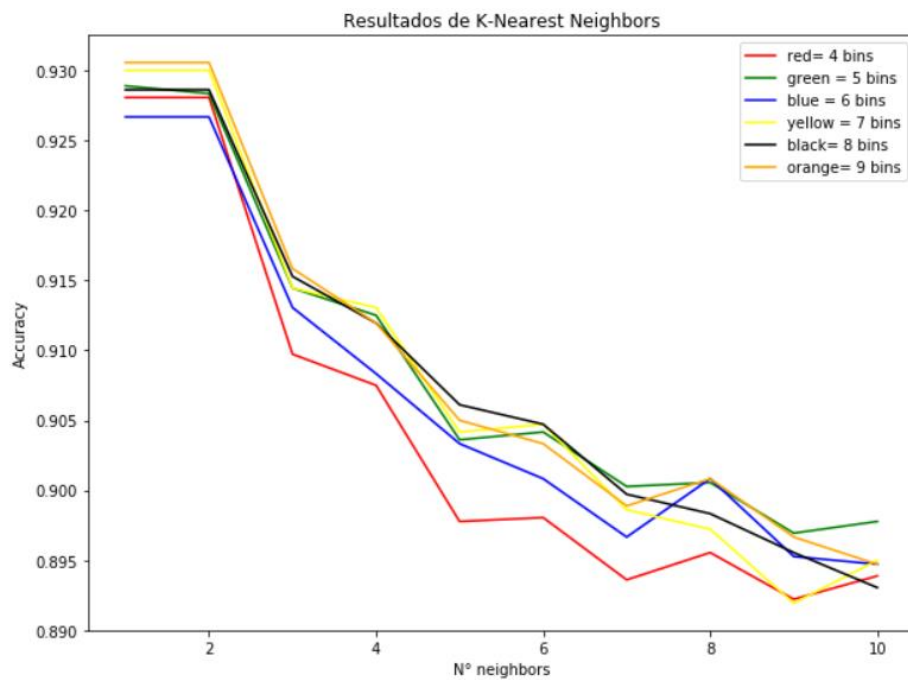


Figura 26: Resultados de K-Nearest Neighbors utilizando HOGs de 4 a 9 bins

Fuente: Elaboraci3n propia

En la Figura 26 se puede observar que los modelos obtienen un mayor accuracy cuando se usa 1 o 2 neighbors y, al aumentar este valor, el accuracy empieza a disminuir. Además, se puede observar que, a pesar de que un HOG de 9 bins obtiene el mejor resultado con 2 neighbors, al aumentar este último valor, el accuracy se reduce hasta ser menor a un HOG de 5 bins, lo cual nos dice que el accuracy obtenido por los modelos que utilicen más de 2 neighbors no siempre mejorara si se utiliza un HOG con altos bins.

El mejor modelo de este algoritmo obtuvo un accuracy de 93.06%, el cual se logró utilizando un HOG de 9 bins y 2 neighbors. En la Figura 27 se puede observar la matriz de confusión de este modelo, donde se observa que las letras con mayor índice de confusión fueron: “m”, “n”, “o”.

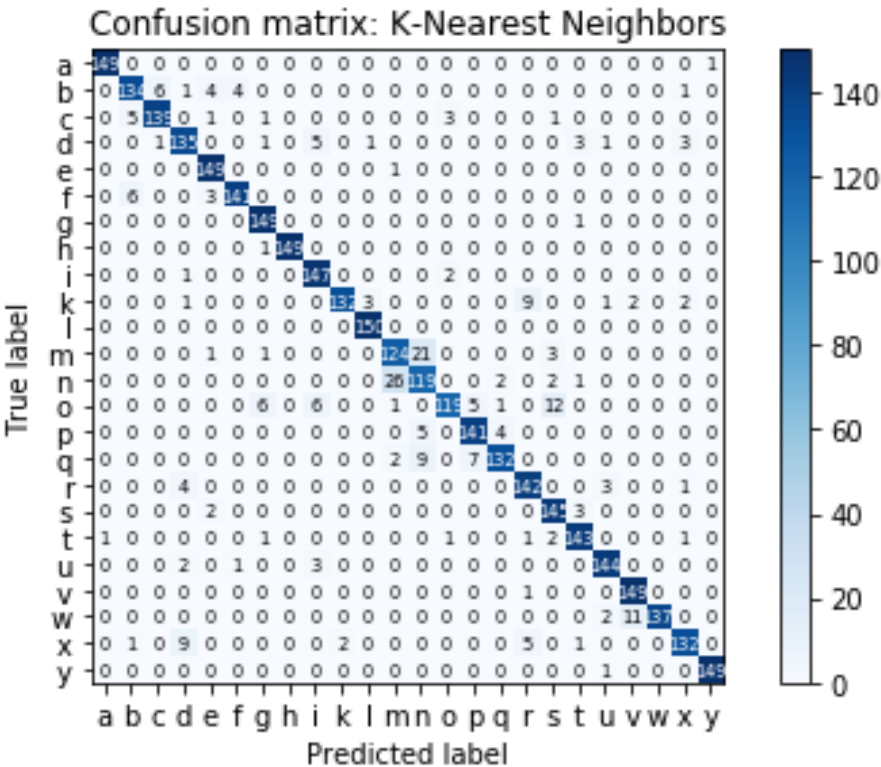


Figura 27: Matriz de confusión de KNN con un HOG de 9 bins

Fuente: Elaboración propia

7.1.3 MultiLayer Perceptron

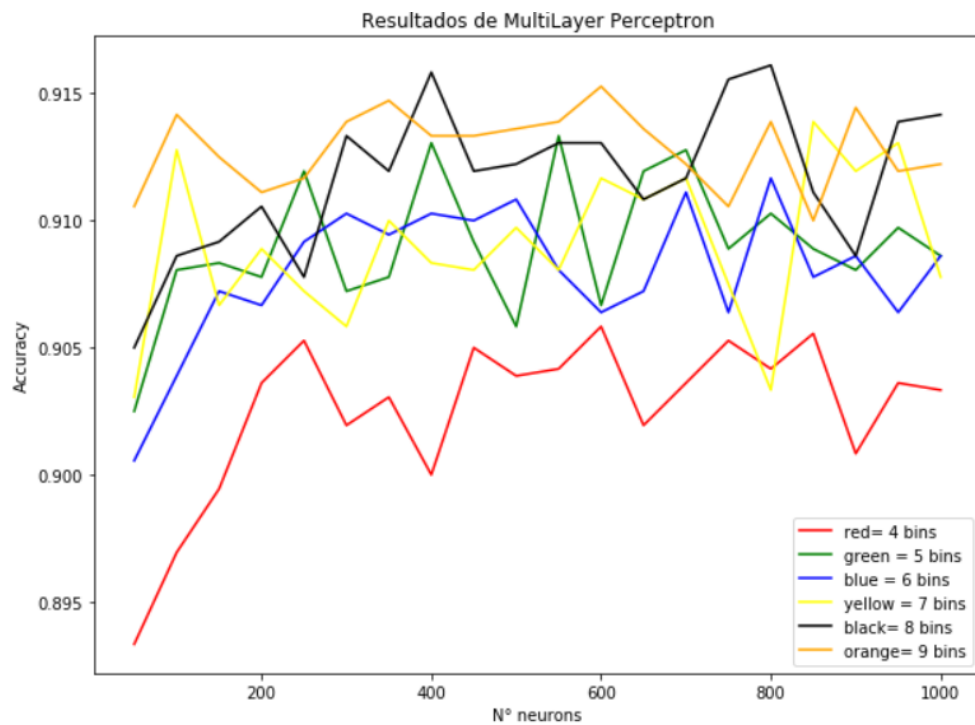


Figura 28: Resultados de MultiLayer Perceptron utilizando HOGs de 4 a 9 bins

Fuente: Elaboración propia

En la Figura 28 se puede observar que existe una dispersión de 1% entre los accuracy obtenidos por los modelos que utilizaron un HOG de 5 a 9 bins, a diferencia del HOG con 4 bins, el cual obtuvo un menor desempeño en casi toda la experimentación.

El mejor modelo de este algoritmo obtuvo un accuracy de 91.61%, el cual se logró utilizando un HOG de 8 bins y 1 hidden layer con 800 neuronas. En la Figura 29 se puede observar la matriz de confusión de este modelo, donde se observa que las letras con mayor índice de confusión fueron: “k”, “m”, “n”, “p”, “q”, “x”.

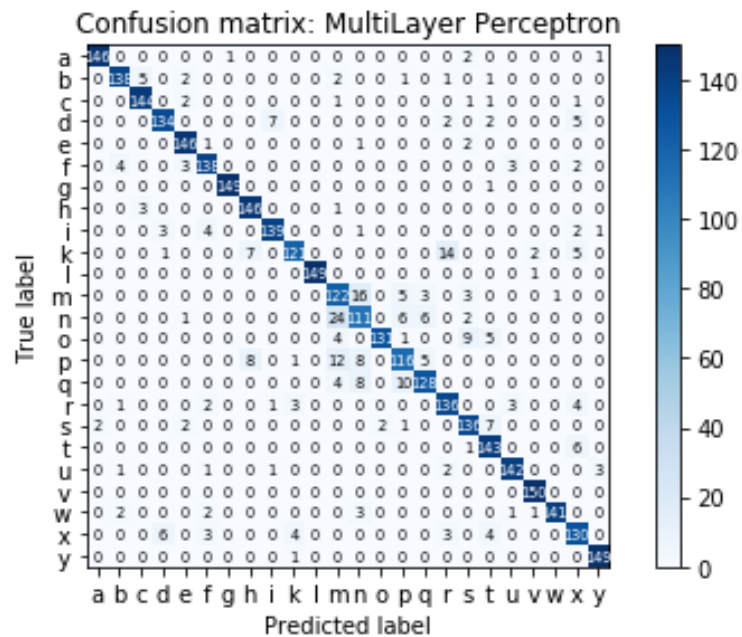


Figura 29: Matriz de confusion de MultiLayer Perceptron utilizando un HOG de 8 bins

Fuente: Elaboración propia

7.1.4 Random Forest

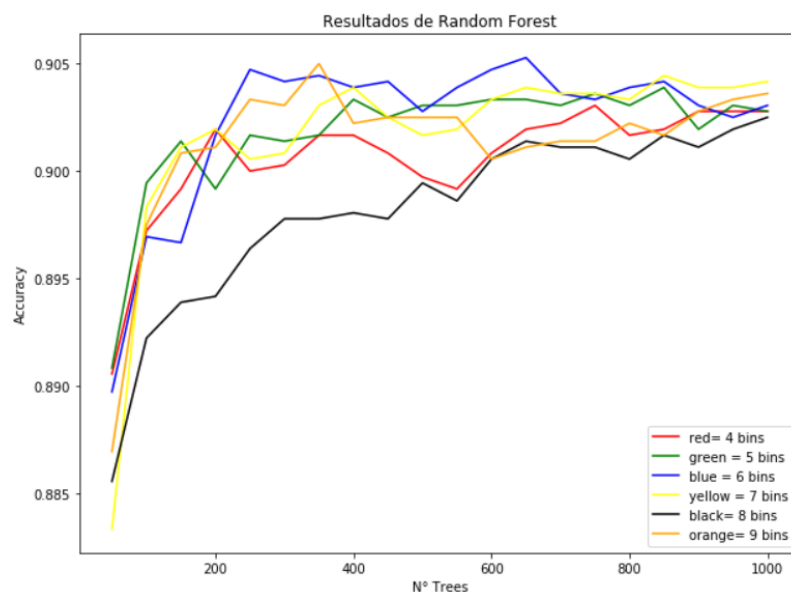


Figura 30: Resultados de Random Forest utilizando HOGs de 4 a 9 bins

Fuente: Elaboración propia

En la Figura 30 se puede observar que los modelos aumentan su accuracy a medida que son entrenados con más árboles. No obstante, a partir de los 200 árboles el accuracy varía según

la cantidad de bins que se use en el HOG, lo cual genera una dispersión de aproximadamente 1%. Sin embargo, esta dispersión tiende a reducirse cuando los modelos se entrenan con 1000 árboles.

El mejor modelo de este algoritmo obtuvo un accuracy de 90.53%, el cual se logró utilizando un HOG de 6 bins y 650 árboles. En la Figura 31 se puede observar la matriz de confusión de este modelo, donde se observa que las letras con mayor índice de confusión fueron: “d”, “k”, “m”, “n”, “o”, “p”, “q”.

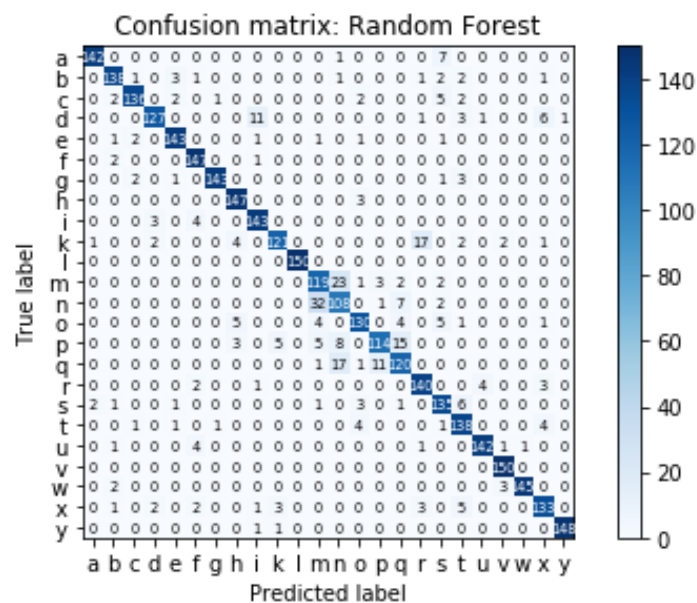


Figura 31: Matriz de confusión de Random Forest utilizando un HOG de 6 bins

Fuente: Elaboración propia

7.1.5 Conclusiones

Los mejores resultados por cada algoritmo se pueden observar en la Tabla 14:

Tabla 14: Mejores resultados obtenidos por cada algoritmo en términos de accuracy

Algoritmo	Bins	Accuracy	Parámetros
Support Vector Machine	9	93.16%	Kernel: rbf, C:1000
K-Nearest Neighbors	9	93.06%	N° neighbors: 2, métrica: manhattan
MultiLayer Perceptron	8	91.61%	hidden layer: 1, N° neuronas: 800
Random Forest	6	90.53%	N° arboles:650, Max-depth: 1000

Fuente: Elaboración propia

En la Tabla 14 se puede observar que los algoritmos de SVM y KNN obtuvieron un mejor accuracy en contraste con los algoritmos de *Random Forest* y *MultiLayer Perceptron*. De los 2 primeros algoritmos, se puede observar en la Figura 1 y Figura 2 que SVM obtuvo mejores resultados en casi todas las pruebas a excepción del HOG con 4 bins, donde el KNN obtuvo un accuracy de 92.81% en comparación al SVM, el cual fue de 92.08%. Una posible causa de esto, se debería a que el KNN logra obtener una mejor distribución de los datos a diferencia de SVM, lo cual haría que KNN obtenga un mejor accuracy.

7.2 Aplicación del modelo de reconocimiento de gestos

Se implementó el modelo de reconocimiento de gestos dentro de un sistema que permite clasificar los 24 gestos estáticos del abecedario de la lengua de señas peruana. En este sistema el usuario deberá adivinar una palabra en base a la cantidad de espacios que aparecen en pantalla. Para ello, deberá levantar su mano dentro del recuadro verde y realizar los gestos del abecedario de la lengua de señas peruana, con el fin de que pueda armar la palabra. La figura 32 muestra el modelo aplicado dentro del sistema.

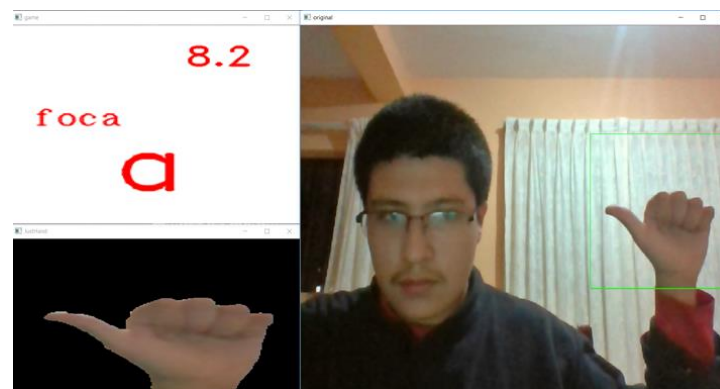


Figura 32: Reconocimiento de la letra “a” utilizando el modelo de clasificación de SVM

Fuente: Elaboración Propia

Las funcionalidades principales de este programa son:

- El programa permite el reconocimiento de los 24 gestos estáticos del abecedario de la lengua de señas peruana.
- El programa permite reconocer gestos con variaciones de traslación, rotación y escala.
- El programa logra la clasificación de los gestos en tiempo real.

Las limitaciones del programa son:

- El programa solo puede identificar gestos estáticos, es decir, no considera gestos que incluyen un movimiento de la mano (gestos dinámicos).
- El programa de reconocimiento de gestos utiliza un algoritmo de sustracción de fondo, donde se requiere que el fondo sea: estático, uniforme y se realice dentro de un ambiente con buena iluminación.

8 GESTIÓN DEL PROYECTO

El presente capítulo describirá los resultados obtenidos en la experimentación para generar el modelo de clasificación.

8.1 Gestión de tiempo

La gestión del tiempo se realizó a través del monitoreo y control de las actividades semanales, las cuales corresponden a los 2 ciclos académicos que duro el desarrollo del proyecto. El proyecto se dividió en 4 fases, de las cuales 2 se desarrollaron en el 1° ciclo y las otras 2 en el 2° ciclo. Estas fases fueron:

- Investigación
- Diseño
- Desarrollo
- Pruebas

Para el 1° ciclo, el cronograma que se desarrollo fue el de la siguiente tabla:

Tabla 15: Cronograma del Proyecto para el 1° ciclo académico

EDT	Modo de tarea	Task Name	Duración	Comienzo	Fin	Predecesoras	Nombres de los recursos
1		Videojuego con reconocimiento de gestos en tiempo real del abecedario de la lengua de señas peruana utilizando cámaras de baja resolución	73 días	mar 15/08/17	vie 24/11/17		
1.1		Inicio	2.38 días	mar 15/08/17	jue 17/08/17		
1.1.1		Semana 1	2.38 días	mar 15/08/17	jue 17/08/17		
1.1.1.1		Elaboración del tema del proyecto	19 hrs	mar 15/08/17	jue 17/08/17		Jefe de Proyecto
1.2		Planificación	9 días	lun 21/08/17	vie 01/09/17		
1.2.1		Semana 2	5 días	lun 21/08/17	vie 25/08/17		
1.2.1.1		Elaboración del Project Charter	40 hrs	lun 21/08/17	vie 25/08/17	4	Jefe de Proyecto
1.2.2		Semana 3	4 días	lun 28/08/17	vie 01/09/17		
1.2.2.1		Investigación de los metodos/técnicas/algoritmos de reconocimiento de gestos de mano en kinect	20 hrs	lun 28/08/17	mié 30/08/17		Jefe de Proyecto
1.2.2.2		Hito: Primera presentación ante el comité	0 hrs	vie 01/09/17	vie 01/09/17	9	Jefe de Proyecto,Jefe de De
1.3		Ejecución	54.13 días	lun 04/09/17	vie 17/11/17		
1.3.1		Semana 4	2.5 días	lun 04/09/17	mié 06/09/17		
1.3.1.1		Investigación sobre los algoritmos de clasificación de reconocimiento de gestos de mano en kinect	20 hrs	lun 04/09/17	mié 06/09/17	10	Jefe de Proyecto
1.3.2		Semana 5	2.5 días	lun 11/09/17	mié 13/09/17		
1.3.2.1		Investigación del dataset del lenguaje de señas americano en kinect	20 hrs	lun 11/09/17	mié 13/09/17	13	Jefe de Proyecto
1.3.2.2		Preparación de la exposición parcial	2 hrs	lun 11/09/17	lun 11/09/17	13	Jefe de Proyecto,Jefe de De
1.3.3		Semana 6	1.25 días	lun 18/09/17	mar 19/09/17		
1.3.3.1		Avance del Objetivo Especifico 1	10 hrs	lun 18/09/17	mar 19/09/17	15	Jefe de Proyecto
1.3.3.2		Hito: Presentación parcial al comité	0 días	mar 19/09/17	mar 19/09/17	15	Jefe de Proyecto,Jefe de De
1.3.4		Semana 7	3.13 días	lun 25/09/17	jue 28/09/17		
1.3.4.1		Solicitud de cambio de camaras de profundidad a camaras de baja resolución	3 hrs	lun 25/09/17	lun 25/09/17	19	Jefe de Proyecto
1.3.4.2		Investigación de metodos/tecnicas/algoritmos de reconocimiento de gestos de mano con camaras de baja resolución	25 hrs	lun 25/09/17	jue 28/09/17	19	Jefe de Proyecto
1.3.5		Semana 9	5.38 días	lun 09/10/17	lun 16/10/17		
1.3.5.1		Reestructurar el Project Charter	3 hrs	lun 09/10/17	lun 09/10/17	21	Jefe de Proyecto
1.3.5.2		Avance del capítulo 3 de la memoria	40 hrs	lun 09/10/17	lun 16/10/17	24	Jefe de Proyecto
1.3.5.3		Implementación del algoritmo de extracción de características	12 hrs	lun 09/10/17	mar 10/10/17	22	Jefe de Proyecto
1.3.5.4		Documentos de Gestión	40 hrs	lun 09/10/17	vie 13/10/17		Jefe de Proyecto
1.3.6		Semana 10	2.5 días	lun 16/10/17	mié 18/10/17		
1.3.6.1		Avance del capítulo 4 de la memoria	10 hrs	lun 16/10/17	mar 17/10/17	25	Jefe de Proyecto
1.3.6.2		Pruebas de concepto del algoritmo implementado	20 hrs	lun 16/10/17	mié 18/10/17	26	Jefe de Proyecto
1.3.7		Semana 11	0.63 días	mar 24/10/17	mar 24/10/17		
1.3.7.1		Avance del Objetivo Especifico 2	5 hrs	mar 24/10/17	mar 24/10/17	30	Jefe de Proyecto
1.3.8		Semana 12	1.88 días	lun 30/10/17	mar 31/10/17		
1.3.8.1		Corrección del capítulo 3	15 hrs	lun 30/10/17	mar 31/10/17	24	Jefe de Proyecto
1.3.9		Semana 13	1 día	vie 10/11/17	vie 10/11/17		
1.3.9.1		Implementación del clasificador en scikit-learn	8 hrs	vie 10/11/17	vie 10/11/17	30	Jefe de Proyecto
1.3.9.2		Presentación de los objetivos especificos 1 y 2	1 hr	vie 10/11/17	vie 10/11/17		Jefe de Proyecto
1.3.10		Semana 14	4.13 días	lun 13/11/17	vie 17/11/17		
1.3.10.1		Avance del capítulo 7 de la memoria	5 hrs	lun 13/11/17	lun 13/11/17	34	Jefe de Proyecto
1.3.10.2		Entrega del DVD del proyecto	1 hr	vie 17/11/17	vie 17/11/17	27	Jefe de Proyecto
1.4		Cierre	0 días	vie 24/11/17	vie 24/11/17		
1.4.1		Semana 15	0 días	vie 24/11/17	vie 24/11/17		
1.4.1.1		Hito: Exposición final con el comité	0 días	vie 24/11/17	vie 24/11/17	11	Jefe de Proyecto, Jefe de Desarrollo

Fuente: Elaboración propia

En la tabla 15, se puede observar que en un inicio se planteó elaborar la propuesta del proyecto en la 1° semana, para después elaborar el Project chárter y empezar la investigación del tema propuesta para la semana 3, debido a que a fines de esa semana se contaba con la 1° exposición ante el jurado, con el fin de exponer un primer avance de la propuesta del proyecto. A partir de la semana 4 hasta la semana 6 se continuó investigando acerca del tema propuesto en las primeras semanas. No obstante, después de la exposición parcial se detectó que el alcance del proyecto sería muy grande para desarrollarlo dentro del tiempo planteado, por lo que se utilizaron las semanas 7 y 9 para replantear el nuevo alcance del proyecto y actualizar el chárter. El nuevo alcance que se planteo fue: “Reconocer gestos de mano estáticos del abecedario de la lengua de señas peruana utilizando cámaras de baja resolución”, en contraste con el antiguo alcance: “Reconocimiento de gestos de mano estáticos y dinámicos del abecedario de la lengua de señas peruana utilizando Kinect”. A partir de la semana 10 hasta la semana 15 se avanzó la etapa del diseño de la metodología a

implementar para la detección de los gestos de mano estáticos, el cual consistió en hacer un diagrama de flujo de la metodología, así como unas pruebas de concepto utilizando un dataset público junto con el HOG los algoritmos de clasificación seleccionados en la investigación. Además, también se contactó con la asociación ENSEÑAS PERÚ y logro llegar a un acuerdo donde ellos nos proporcionarían a los expertos en señas para hacer el dataset a cambio de darles el programa de reconocimiento de gestos dentro de un juego educativo.

Para el 2º ciclo, el cronograma que se desarrollo fue el de la siguiente tabla:

Tabla 16: Cronograma del proyecto para el 2º ciclo académico

Fecha	DESCRIPCIÓN DE LAS ACTIVIDADES	Porcentaje de Avance (%)	Responsable	ENTREGABLES
Semana 0	Reunión con el Cliente del proyecto	100	Bryan Berrú	1º Acta de reunión de Cliente
Semana 1	Elaboración del plan de trabajo	100	Bryan Berrú	
Semana 1	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 1	Avance del Objetivo 3 - Investigación de Descriptores de Fondo	100	Bryan Berrú	
Semana 1	Reunión con el Cliente del proyecto	100	Bryan Berrú	2º Acta de reunión de cliente
Semana 1	Avance del Objetivo 3 - Implementación del BackgroundSubtractor	100	Bryan Berrú	
Semana 1	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 2	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 2	Reunión con el Cliente del proyecto	100	Bryan Berrú	3º Acta de reunión de Cliente
Semana 2	Avance del Objetivo 3 - 1ª Visita a ENSEÑAS PERU para la recolección de gestos	100	Bryan Berrú	
Semana 2	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 2	Avance del Objetivo 4 - Generación del T1 modelo de clasificación	100	Bryan Berrú	
Semana 3	Reunión con el Portfolio Manager	100	Bryan Berrú	4º Acta de reunión del Cliente
Semana 3	Reunión con el Cliente	100	Bryan Berrú	
Semana 3	Avance del Objetivo 3 - 2ª Visita a ENSEÑAS PERU para la recolección de gestos	100	Bryan Berrú	
Semana 3	Avance del Objetivo 3 - Etiquetado manual de los gestos recolectados (4 personas)	100	Bryan Berrú	
Semana 3	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 3	Solicitud de Recurso de QS	100	Bryan Berrú	
Semana 4	Reunión con el Cliente del proyecto	100	Bryan Berrú	5º Acta de reunión de Cliente
Semana 4	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 4	Avance del Objetivo 3 - 3ª Visita a ENSEÑAS PERU para la recolección de gestos	100	Bryan Berrú	
Semana 4	Avance del Objetivo 3 - Etiquetado manual de los gestos recolectados (6 personas)	100	Bryan Berrú	
Semana 4	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 4	Avance del Objetivo 3 - Implementación del Hand Tracking	100	Bryan Berrú	
Semana 4	Revisión de avance por parte de QS del Sprint 4	100	Bryan Berrú y Recurso de QS	Acta de Conformidad del recurso de QS
Semana 5	Reunión con el Cliente del proyecto	100	Bryan Berrú	6º Acta de reunión de Cliente
Semana 5	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 5	Avance del Objetivo 3 - 4ª Visita a ENSEÑAS PERU para la recolección de gestos	100	Bryan Berrú	
Semana 5	Avance del Objetivo 4 - Implementación de los algoritmos de scikit-learn	100	Bryan Berrú	
Semana 5	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 5	Avance del Capítulo 5 de la Memoria - Desarrollo del Proyecto	100	Bryan Berrú	
Semana 5	Solicitud de Recurso para el desarrollo del juego	100	Bryan Berrú	
Semana 5	Avance de la Cartera de Proyectos	100	Bryan Berrú	
Semana 6	Reunión con el Cliente del proyecto	100	Bryan Berrú	7º Acta de reunión de Cliente
Semana 6	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 6	Avance del Objetivo 3 - Etiquetado manual de los gestos recolectados (5 personas)	100	Bryan Berrú	
Semana 6	Avance del Objetivo 4 - Conectividad de Unity y Python	100	Recurso de SF	
Semana 6	Avance del Objetivo 4 - Generación del 2º modelo de clasificación	100	Bryan Berrú	
Semana 6	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 6	Avance del Capítulo 5 de la Memoria - Desarrollo del Proyecto	100	Bryan Berrú	
Semana 6	Avance de Propuesta de Cartera de Proyectos	100	Bryan Berrú	
Semana 6	Validación del Objetivo Especifico 3 por parte de ENSEÑAS PERU	100	Bryan Berrú	Acta de Conformidad del Experto para el Objetivo 3
Semana 7	Reunión con el Cliente del proyecto	100	Bryan Berrú	8º Acta de reunión de Cliente
Semana 7	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 7	Avance de Propuesta de Cartera de Proyectos	100	Bryan Berrú	
Semana 7	Avance del Objetivo 4 - Desarrollo del videojuego en Unity	100	Recurso de SF	
Semana 7	Avance del Capítulo 5 de la Memoria - Desarrollo del Proyecto	100	Bryan Berrú	
Semana 7	Finalización y Validación del Objetivo Especifico 3 por parte del Cliente	100	Bryan Berrú	Acta de Conformidad del Cliente para el Objetivo 3
Semana 7	Solicitud de recuso de QS	100	Bryan Berrú	
Semana 7	Reunión con el Portfolio Manager	100	Bryan Berrú	
Semana 8	Sustentación Parcial con el Portfolio Manager	100	Bryan Berrú	

Semana 9	Reunión con el Cliente del proyecto	100	Bryan Berrú	9ª Acta de reunión de Cliente
Semana 9	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 9	Entrega de la Propuesta de Cartera de Proyectos	100	Bryan Berrú	Documento de cartera de Proyectos
Semana 9	Finalización del Capítulo 5 de la Memoria - Desarrollo del Proyecto	100	Bryan Berrú	Capítulo 5 de la memoria
Semana 9	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 9	Avance del Objetivo 4 - Experimentación con el dataset recolectado con los 4 algoritmos de clasificación	100	Bryan Berrú	
Semana 9	Revisión de avance por parte de QS del Sprint 6	100	Bryan Berrú	Acta de conformidad del recurso de QS
Semana 9	Sustentación Parcial con el Comité	100	Bryan Berrú	
Semana 10	Reunión con el Cliente del proyecto	100	Bryan Berrú	10ª Acta de reunión de Cliente
Semana 10	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 10	Avance del Capítulo 6 de la Memoria - Resultados del Proyecto	100	Bryan Berrú	
Semana 10	Avance del Objetivo 4 - Evaluación de los experimentos	100	Bryan Berrú	
Semana 10	Avance del Objetivo 4 - Aplicación del modelo dentro del videojuego	100	Recurso de SF	
Semana 10	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 11	Reunión con el Cliente del proyecto	100	Bryan Berrú	11ª Acta de reunión de Cliente
Semana 11	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 11	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 11	Finalización y Validación del Objetivo 4	100	Bryan Berrú	Acta de Conformidad del Cliente para el Objetivo 4
Semana 11	Avance del Capítulo 6 de la Memoria - Resultados del Proyecto	100	Bryan Berrú	
Semana 11	Avance del Capítulo 2 de la Memoria - Student Outcomes	100	Bryan Berrú	
Semana 11	Avance del Anexo C - Costos y Presupuestos	100	Bryan Berrú	
Semana 12	Reunión con el Cliente del proyecto	100	Bryan Berrú	12ª Acta de reunión de Cliente
Semana 12	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 12	Finalización del Paper	100	Bryan Berrú	Artículo científico
Semana 12	Avance del Capítulo 2 de la Memoria - Student Outcomes	100	Bryan Berrú	
Semana 12	Finalización del Capítulo 6 de la Memoria - Resultados del Proyecto	100	Bryan Berrú	Capítulo 6 de la memoria
Semana 12	Finalización del Anexo C - Costos y Presupuestos	100	Bryan Berrú	Anexo C de la memoria
Semana 12	Avance del Anexo A - WASC	100	Bryan Berrú	
Semana 12	Solicitud de Recurso de QS	100	Bryan Berrú	
Semana 12	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 13	Reunión con el Cliente del proyecto	100	Bryan Berrú	13ª Acta de reunión de Cliente
Semana 13	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 13	Finalización del Capítulo 2 de la Memoria - Student Outcomes	100	Bryan Berrú	Capítulo 2 de la memoria
Semana 13	Finalización Anexo A - WASC	100	Bryan Berrú	Anexo A de la memoria
Semana 13	Revisión por parte de QS para el sprint 7	100	Bryan Berrú y Recurso de QS	Acta de Conformidad del recurso de QS
Semana 13	Avance del anexo D - Análisis de Complejidad Algorítmica	100	Bryan Berrú	
Semana 13	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 14	Reunión con el Cliente del proyecto	100	Bryan Berrú	14ª Acta de reunión de Cliente
Semana 14	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 14	Avance del anexo D - Análisis de Complejidad Algorítmica	100	Bryan Berrú	
Semana 14	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 14	Desarrollo de Conclusiones y Recomendaciones del proyecto	100	Bryan Berrú	Conclusiones y Recomendaciones
Semana 15	Reunión con el Cliente del proyecto	100	Bryan Berrú	15ª Acta de reunión de Cliente
Semana 15	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 15	Finalización del Anexo D - Análisis de Complejidad Algorítmica	100	Bryan Berrú	Anexo D de la memoria
Semana 15	Reunión con el Portafolio Manager	100	Bryan Berrú	
Semana 15	Sustentación Final con el Portafolio Manager	100	Bryan Berrú	
Semana 15	Preparación y Entrega de los artefactos de toda la memoria	100	Bryan Berrú	Artefactos de la Memoria
Semana 16	Sustentación Final con el Comité	0	Bryan Berrú	

Fuente: Elaboración propia

En la tabla 16, se puede observar que las primeras 7 semanas se enfocaron en desarrollar los algoritmos de HOG y Hand Tracking, con el fin de detectar la región de la mano y obtener el vector de características. Adicionalmente, cada semana, se realizó una visita a la asociación ENSEÑAS PERU, con el fin de recolectar el dataset de imágenes de los gestos estáticos del abecedario de la lengua de señas peruana. Una vez obtenido todo el dataset de imágenes para la semana 9, se empezó a realizar diversas pruebas utilizando diferentes parámetros en el HOG y en los 4 algoritmos de clasificación. Estas pruebas luego fueron registradas dentro de la memoria y el paper. Finalmente, en las últimas 3 semanas, se completó los últimos anexos y capítulos de la memoria.

8.2 Gestión de las Comunicaciones

Comunicación con el Profesor Cliente: Se emplearon dos canales de comunicación con el cliente: reuniones presenciales y virtuales (HangOut, Skype, Whatsapp). Las reuniones presenciales se llevaron una o dos veces a la semana durante el ciclo académico, las cuales se conformaron por las siguientes actividades:

- Revisión del progreso del proyecto
- Aprobación de los indicadores de éxito, cronograma, Project charter y memoria
- Consultas referentes a los algoritmos a implementar a lo largo del proyecto

Sobre las reuniones virtuales, estas se llevaron una vez a la semana y se empleó principalmente para mostrar los avances de cada uno de los objetivos, además de facilitar la resolución de dudas que se hayan presentado a lo largo de la semana o que no fueron resueltas en una reunión presencial.

Todas las reuniones presenciales y virtuales fueron acompañadas de un acta de reunión como un medio sostenible para mantener los acuerdos dados en cada una de las reuniones.

Comunicación con el Profesor Gerente de InnovaTi: Las reuniones se realizaron presencialmente, cada semana, dentro de las horas de taller de proyecto. Estas reuniones involucraron las siguientes actividades:

- Revisión del progreso del proyecto
- Aprobación de los documentos de gestión, cronograma, Project charter y memoria

Todas las reuniones presenciales fueron acompañadas de un acta de reunión como un medio sostenible para mantener los acuerdos dados en cada una de las reuniones.

Comunicación con la organización ENSEÑAS PERÚ: Se realizó un total de 2 reuniones presenciales con ENSEÑAS PERÚ las cuales fueron presenciales y se realizaron en la semana 11 y 13 del ciclo académico. En la primera reunión, se presentó el proyecto a la organización y la segunda reunión fue para firmar el acta de reunión donde se establece los acuerdos llegados por ambos lados.

8.3 Gestión de los Riesgos

Los riesgos planteados para este proyecto son los siguientes:

Tabla 17: Matriz de Riesgos

Cód. Riesgo	Fecha de registro	Categoría - Origen	Descripción del Riesgo	Disparador	Probabilidad	Impacto (Ht)	Impacto
20172-001-R001	25/10/2017	Int - Equipo de trabajo	Cambio de cliente asignado	El cliente se retira de la UPC por motivos mayores o el comité	30% Baja	25	5 Muy alta
20172-001-R002	25/10/2017	Int - El proyecto	Dificultad con la implementación del de la metodología a utilizar	Falta de conocimiento en el área	30% Baja	10	2 Baja
20172-001-R003	25/10/2017	Ext - Usuarios	Dificultad para conseguir el apoyo de la asociación de sordos	Falta de interés o tiempo de la asociación de sordos	10% Muy baja	25	5 Muy alta
20172-001-R004	25/10/2017	Int - Equipo de trabajo	Retiro del jefe de proyecto o jefe de desarrollo	El comité no aprueba el trabajo realizado por el jefe de proyecto o de desarrollo	10% Muy baja	30	5 Muy alta

Fuente: Elaboración propia

En el transcurso de este ciclo no se ha presentado ninguno de los riesgos mencionados en la matriz de riesgos. No obstante, se han planteado planes de contingencia en caso uno de estos riesgos se manifieste.

8.4 Lecciones Aprendidas

A lo largo del ciclo, sobre todo en la etapa de planificación se ha podido observar que para definir el tema y el alcance del proyecto se debe considerar primeramente que los objetivos que se proponen en el proyecto deben cumplir con el formato SMART, lo cual conlleva a que cada objetivo sea analizado de manera que pueda ser factible de realizar dentro de los 2 ciclos académicos de taller de proyectos. Para ello, es necesario recopilar la mayor cantidad de información posible y analizar los diversos enfoques que tiene el tema a tratar, de manera que podamos tener una mejor perspectiva de la situación y poder presentar un aporte considerable a la carrera.

9 CONCLUSIONES

A continuación, se presentan las conclusiones obtenidas en la elaboración del proyecto para cada una de sus fases. Además, se culminará proponiendo futuros proyectos.

En este proyecto se aplicó un modelo de clasificación de reconocimiento de gestos estáticos del abecedario de la lengua de señas peruana, utilizando cámaras de baja resolución, dentro de un sistema de reconocimiento, que permite aprender, de manera interactiva, los gestos de la lengua de señas peruana, mediante el deletreo de palabras. Para cumplir este objetivo, se dividió el proyecto en 4 fases.

1. En la primera fase, se investigó diversos artículos científicos, con el fin de determinar la metodología más eficiente para el desarrollo del proyecto. Dentro de esta investigación se concluyó que los 4 algoritmos más relevantes para realizar el reconocimiento de los gestos de mano fueron: Support Vector Machine, K-Nearest neighbor, MultiLayer Perceptron y Random Forest. Además, estos algoritmos lograron obtener un mejor accuracy utilizando el descriptor de imagen de Histogram Oriented Gradient, el cual se utilizó también en el proyecto.

2. En la segunda fase, se diseñó el diagrama de flujo de actividades de los componentes del sistema de reconocimiento de gestos. Para ello, esta etapa se dividió en 3 sub-fases, la primera se enfocó en la recolección del dataset. La segunda, en el pre-procesamiento y la extracción de las características de las imágenes. Finalmente, la tercera se enfocó en la generación del modelo de clasificación y su respectiva integración dentro del sistema de reconocimiento. En base a este diseño se concluyó que el sistema es capaz de: reconocer los 24 gestos estáticos del abecedario de la lengua de señas peruana y lograr una clasificación de los gestos en tiempo real.

3. En la tercera fase, se construyó el dataset de imágenes, de los gestos estáticos del abecedario de la lengua de señas peruana. Debido a que este dataset estaba relacionado a una lengua existente, fue necesario trabajar junto a personas sordas, de manera que puedan ayudarnos a generar y validar cada una de las imágenes que se utilizaran dentro de este dataset. Para ello, se trabajó junto con la asociación ENSEÑAS PERU, la cual proveyó de personas sordas que permitieran realizar cada una de los 24 gestos a utilizar en el dataset,

además de validar las imágenes. Gracias a ello, se logró crear un dataset robusto, que incluye imágenes con variaciones de rotación y escala, además de contar con gestos realizados por expertos.

4. En la cuarta fase, se realizó la experimentación del dataset utilizando los 4 algoritmos de clasificación seleccionados en el objetivo específico 1. De esta fase, se concluyó que el algoritmo que consigue el mejor resultado, en términos de accuracy, fue el SVM, con un 93.16% utilizando un HOG de 9 bins.

10 RECOMENDACIONES

A continuación, se dará una serie de recomendaciones para el desarrollo de futuros proyectos de tesis, relacionados al desarrollo de reconocimiento de gestos de mano.

- Se recomienda expandir la cantidad de gestos a reconocer sobre la lengua de señas peruana, de manera que se pueda reconocer gestos dinámicos. Esto daría pase, a que se pueda crear un videojuego más completo, con el fin de facilitar la enseñanza de la lengua de señas peruana.
- Se recomienda que este dataset pueda ser utilizado junto con algoritmos de Deep, de manera que se pueda expandir el análisis algorítmico del modelo de clasificación, mediante la comparación de resultados entre diferentes arquitecturas de redes neuronales y los resultados obtenidos con los algoritmos de clasificación de Machine Learning que se utilizaron en este proyecto.
- Se recomienda que se utilice el dataset actual, con el fin de crear un videojuego más inmersivo utilizando leap motion y VR.

11 GLOSARIO

A continuación, se presenta un glosario de los términos utilizados en el proyecto:

- OpenCV: Librería multiplataforma de procesamiento de imágenes
- Scikit-learn: Librería de machine learning en python que permite utilizar una serie de algoritmos de clasificación para generar un modelo.
- HOG: Histogram Oriented Gradient
- SVM: Support Vector Machine
- KNN: K-Nearest Neighbors
- MLP: MultiLayer Perceptron
- RF: Random Forest

12 REFERENCIAS

Ruchi Manish Gurav, Premanand K. Kadbe. (2015). Real time Finger Tracking and Contour Detection for Gesture Recognition using OpenCV. International Conference on Industrial Instrumentation and Control, 974-977. 02/10/2017, De IEEE Base de datos.

Mendoza D., Barreno M., Velasco N.. (2017). Novel Method for using Hand Recognition as Computer Remote Control with Computer Vision Techniques. Journal of Telecommunication Electronic and Computer Engineering, 9, 89-92. 28/10/2017, De UTEM Base de datos.

ABD Albary Sulyman, Zeyad T. Sharef, Kamaran Hama Ali Faraj, Zaid Ahmed Aljawaryy, Fahad Layth Malallah . (2017). Real-Time Numerical 0-5 Counting Based on Hand-Finger Gestures Recognition. Journal of Theoretical and Applied Information Technology, 95, 3105-3115. 25/10/2017, De Scopus Base de datos.

Lan Tiantian, Shen Jinyuan, Liu Runjie, Guo Yingying . (2015). Hand Gesture Recognition Based on Improved Histograms of Oriented Gradients. 27th Chinese Control and Decision Conference, 4211-4215. 28/10/2017, De IEEE Base de datos.

X. B. Meng J. Lin. An extended HOG model: SCHOG for human hand detection, Proceedings of 2012 International Conference on Systems and Informatics, Yantai, IEEE, 2593 -2596, 2012.

Jayshree R. Pansare, Maya Ingle. (2016). Vision-Based Approach for American Sign Language Recognition Using Edge Orientation Histogram. International Conference on Image, Vision and Computing , 86-90. 02/10/2017, De IEEE Base de datos.

Amit Kumar Gautam. (2017). American Sign Language Recognition System Using Image Processing Method. International Journal on Computer Science and Engineering , 9, 466-471. 29/10/2017, De ENNGJournarls Base de datos.

Thao Nguyen Thi Huong, Tien Vu Huu, Thanh Le Xuan, San Vu Van . (2015). Static hand gesture recognition for Vietnamese sign language (VSL) using principle components analysis. International Conference on Communications, Management and Telecommunications, 138-141. 28/10/2017, De IEEE Base de datos.

Md. Hafizur Rahman, Jinia Afrin. (2013). Hand Gesture Recognition using Multiclass Support Vector Machine. International Journal of Computer Applications , 74, 39-43. 29/10/2017, De IEEE Base de datos.

Gjorgji Strezoski, Dario Stojanovski, Ivica Dimitrovski, Gjorgji Madjarov. (2016). Hand Gesture Recognition using Deep Convolutional Neural Networks. ICT Innovations,1-10. 01/10/2017, De Springer Base de datos.

Sriparna Saha, Rimita Lahiri, Amit Konar, Atulya K. Nagar. (2016). A Novel Approach To American Sign Language Recognition Using MAdaline Neural Network . IEEE Symposium Series on Computational Intelligence, 16, 1-6. 26/09/2017, De IEEE Base de datos.

S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," in Signal Processing and Information Technology, 2007 IEEE International Symposium on, 2007, pp. 11–16.

Vladislava Bobić, Predrag Tadić, Goran Kvaščev . (2016). Hand gesture recognition using neural network based techniques. Symposium on Neural Networks and Applications (NEUREL), 13, 1-4. 29/09/2017, De IEEE Base de datos

Zoran Zivkovic,Ferdinand van der Heijden. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recognition Letters, vol 27. 773-780., Scopus.

Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

13 ANEXOS

Anexo A: Lineamiento WASC

Anexo C: Documento de Costos y Presupuestos

Anexo D: Documento de Análisis de Complejidad Algorítmica